THE ANNALS
of
MATHEMATICAL STATISTICS

VOLUME XVIII

1947

THE ANNALS
OF MATHEMATICAL STATISTICS

CONTENTS OF VOLUME XVIII

ARTICLES

NOTES

MISCELLANEOUS

$p \leq q$, and the sample with � to the
dependent variate is divided in the u...... ...... , ⎦to a part
with q degrees of freedom corresponding to the independent variate and the
remaining part with $n - q$ degrees of freedom.   If a_{ij}, b_{ij} denote the sums of
squares and products corresponding to this division, then it is known that the
joint distribution of a_{ij} and b_{ij}, if the dependent vector variate is normal and
actually, in the statistical sense, independent of the second vector variate, is

$$(1) \qquad \frac{|A|^{\frac{1}{2}(q-p-1)}|B|^{\frac{1}{2}(n-q-p-1)} \exp\left[-\frac{1}{2}\sum_{i=1}^{p}(a_{ii}+b_{ii})\right] da\, db}{2^{\frac{1}{2}np}\,\pi^{\frac{1}{2}p(p-1)}\prod_{i=0}^{p-1}\{\Gamma[\frac{1}{2}(q-i)]\Gamma[\frac{1}{2}(n-q-i)]\}}$$

where $|A|$ denotes the determinant of the matrix $A \equiv \{a_{ij}\}$, and da the product
of differentials da_{ij}, and where for convenience the variance matrix of the
dependent variate is taken to be the unit matrix.

We make the transformation specified by

$$(2) \qquad \begin{aligned} A &= WDW', \\ A + B &= WW', \end{aligned}$$

where D is a diagonal matrix of the quantities r_i^2 in descending order of magnitude,
and $W = \{w_{ij}\}$ is a matrix (with transpose W') uniquely determined by (2)
except for an ambiguity of sign for each column; this ambiguity can be eliminated
by choosing positive elements in the first row.   The Jacobian Δ of the trans-
formation may be shown to be

$$(3) \qquad \Delta = 2^{p}|WW'|^{\frac{1}{2}p+1}\prod_{i=1}^{p}\prod_{j=i+1}^{p}(r_i^2 - r_j^2).$$

By direct substitution, we obtain from (1) the distribution

$$p(a_{ij}, b_{ij}) = p(w_{ij}, r_i^2) = p(w_{ij})p(r_i^2),$$

where $p(x)$ is a general notation[3] for a distribution function in one or more
variates x, (including the differential elements); for $p(w_{ij})$ and $p(r_i^2)$ we have

$$(4) \qquad p(w_{ij}) = C_1|WW'|^{\frac{1}{2}(n-p)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{p} w_{ij}^2\right] dw,$$

$$(5) \qquad p(r_i^2) = C_2\prod_{i=1}^{p}\left\{(r_i^2)^{\frac{1}{2}(q-p-1)}(1-r_i^2)^{\frac{1}{2}(n-q-p-1)}\prod_{j=i+1}^{p}(r_i^2 - r_j^2)\right\} dr^2,$$

---

[3] The probability symbol is not of course to be confused with the number p of components
in the dependent variate.   It should also be noted that for convenience $p(x_i)$ is used to
denote the *joint* probability for a set of quantities x_i, whereas $p(x_1)$ or $p(x_2)$ denotes the
probability for the specified variate x_1 or x_2 considered separately.

the constants C_1 and C_2 being arranged to give unity on integration of $p(w_{ij})$ or $p(r_i^2)$, i.e. we have

(6)
$$C_1 = 2^{p-\frac{1}{2}np}\, \pi^{-\frac{1}{4}p^2} \prod_{i=0}^{p-1} \{\Gamma[\tfrac{1}{2}(p - i)]/\Gamma[\tfrac{1}{2}(n - i)]\},$$

(the w_{ij} varying from $-\infty$ to ∞ except that $w_{1i} \geq 0$), and

(7)
$$C_2 = \pi^{\frac{1}{2}p} \prod_{i=0}^{p-1} \{\Gamma[\tfrac{1}{2}(n - i)]/(\Gamma[\tfrac{1}{2}(p - i)]\Gamma[\tfrac{1}{2}(q - i)]\Gamma[\tfrac{1}{2}(n - q - i)])\}.$$

3. Formal determination of the general distribution. The method to be adopted of obtaining the general distribution from the particular case quoted in equation (5) above is the same in principle as the one adopted by Fisher [7] in his derivation of the general distribution of the multiple correlation coefficient. Since the argument is more involved in the present problem, it will be presented first in formal probability terms, before the details of the solution are examined.

We consider a transformation of the components of each vector variate to the true canonical components. Let the observed ordinary correlation coefficients of these mutually independent components for one vector variate with the corresponding components of the second vector variate be denoted by s_i. The true correlations are the true canonical correlations ρ_i. Then we have for the general canonical correlation distribution denoted by[4] $p(r_i \mid \rho_i)$, the expression

$$p(r_i \mid \rho_i) = \int_{s_i} p(r_i, s_i \mid \rho_i)$$

$$= \int_{s_i} p(r_i \mid s_i, \rho_i)p(s_i \mid \rho_i)$$

$$= \int_{s_i} p(r_i \mid s_i)p(s_1 \mid \rho_1)p(s_2 \mid \rho_2) \cdots p(s_p \mid \rho_p),$$

the substitution $p(r_i \mid s_i)$ for $p(r_i \mid s_i, \rho_i)$ following from the sufficiency of the independent correlations s_i of the corresponding pairs of canonical components, as statistics for the ρ_i. We now define the function $g(s_1, \rho_1)$ by the relation

$$p(s_1 \mid \rho_1) = p(s_1 \mid \rho_1 = 0)\, g(s_1, \rho_1),$$

whence we have the general solution

(8)
$$p(r_i \mid \rho_i) = \int_{s_i} p(r_i \mid s_i)p(s_1 \mid \rho_1 = 0)g(s_1, \rho_1)p(s_2 \mid \rho_2 = 0)g(s_2, \rho_2) \cdots$$

$$= \int_{s_i} p(r_i, s_i \mid \rho_i = 0)g(s_1, \rho_1)g(s_2, \rho_2) \cdots$$

$$= p(r_i \mid \rho_i = 0) \int_{s_i} p(s_i \mid r_i, \rho_i = 0)g(s_1, \rho_1)g(s_2, \rho_2) \cdots$$

for $p(r_i \mid \rho_i)$ in terms of the special case $p(r_i \mid \rho_i = 0)$.

---

[4] Quantities to the right of the vertical stroke in a probability bracket are *given* quantities on which the probability distribution depends.

Now according as the independent vector variate is considered as (a) a normal variate with which the dependent variate is correlated, (b) a fixed vector in sample space (this includes the non-central means case) Fisher [7] has shown that the distribution of the *multiple* correlation R of a single dependent variate with an independent variate comprising m components is $p(R \mid \rho = 0)g(R, \rho)$, where

$$(9) \quad \begin{aligned} \text{(a)} &\qquad g(R, \rho) = F(\tfrac{1}{2} n, \tfrac{1}{2} n; \tfrac{1}{2} m; \rho^2 R^2)\,(1 - \rho^2)^{\frac{1}{2}n}, \\ \text{(b)} &\qquad g(R, \rho) = F(\tfrac{1}{2} n; \tfrac{1}{2} m; \tfrac{1}{2} \beta^2 R^2)\,e^{-\frac{1}{2}\beta^2}, \end{aligned}$$

where we replace ρ^2 by a parameter β^2 in case (b), and the notation for hypergeometric functions used is:

$$F(\alpha; \beta; x) = 1 + \frac{\alpha x}{\beta} + \frac{\alpha(\alpha + 1)x^2}{\beta(\beta + 1)2!} + \cdots,$$

$$F(\alpha_1, \alpha_2 ; \beta; x) = 1 + \frac{\alpha_1 \alpha_2 x}{\beta} + \frac{\alpha_1(\alpha_2 + 1)\alpha_2(\alpha_2 + 1)x^2}{\beta(\beta + 1) 2!} + \cdots.$$

It follows that we may write $g(s_1, \rho_1)$ above in the form

$$(10) \quad \begin{aligned} \text{(a)} &\qquad g(s_1, \rho_1) = F(\tfrac{1}{2} n, \tfrac{1}{2} n; \tfrac{1}{2}; \rho_1^2 s_1^2)\,(1 - \rho_1^2)^{\frac{1}{2}n}, \\ \text{(b)} &\qquad g(s_1, \rho_1) = F(\tfrac{1}{2} n; \tfrac{1}{2}; \tfrac{1}{2}\beta_1^2 s_1^2)e^{-\frac{1}{2}\beta_1^2}, \end{aligned}$$

by putting $m = 1$ in (9), (the signs of the s_i are arbitrary, so that we are essentially concerned, as in the multiple correlation distribution, with the squares of the correlations). From these series expansions the integral in (8) consists of terms corresponding to the conditional moments, for any set of positive integers $t_1, t_2, \cdots, t_p$,

$$\begin{aligned} \mu(t_1, t_2, \cdots, t_p) &\equiv E\{(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p} \mid r_i\} \\ &= \int_{s_i} (s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}\, p(s_i \mid r_i, \rho_i = 0). \end{aligned}$$

In the particular case when only $\rho_1 \neq 0$, the moments $\mu(t) \equiv E\{(s_1^2)^t \mid r_i\}$ from the single factor $g(s_1, \rho_1)$ are all that arise, but in the general case it is important to notice that the quantities s_i^2, while statistically independent when unrestricted, are no longer independent for the *conditional* distribution $p(s_i \mid r_i, \rho_i = 0)$.

This completes the formal solution.   It remains to evaluate $\mu(t_1, t_2, \cdots, t_p)$.

4. The conditional moment $\mu(t_1, t_2, \cdots, t_p)$.   First of all we note from the choice of the components of the dependent vector variate, applying the analysis of section 2 to such components, that the multiple correlation R_i between the ith component and the q components of the independent variate is given by

$$R_i^2 = a_{ii}/(a_{ii} + b_{ii}) = \alpha_{i1}^2 r_1^2 + \alpha_{i2}^2 r_2^2 + \cdots + \alpha_{ip}^2 r_p^2,$$

where

$$\alpha_{ij} = w_{ij}/\sqrt{(w_{i1}^2 + w_{i2}^2 + \cdots + w_{ip}^2)}.$$

To obtain the distribution of the α_{ij} from that of the w_{ij}, we note that the w_{ij} distribution (4) is normal (allowing for convenience w_{1i} to vary from $-\infty$ to ∞) except for the "linkage factor"

$$2^{-\frac{1}{2}p(n-p)} \, |\, WW'\,|^{\frac{1}{2}(n-p)} \prod_{i=0}^{p-1} \{\Gamma[\tfrac{1}{2}(p-i)/\Gamma[\tfrac{1}{2}(n-i)]]\}.$$

Hence if we transform to the variables c_{ii}, θ_{ij} defined by

$$c_{ii} = w_{i1}^2 + w_{i2}^2 + \cdots + w_{ip}^2,$$

$$\alpha_{i1} = \cos\theta_{i1},$$

$$\alpha_{i2} = \sin\theta_{i1}\cos\theta_{i2},$$

(11)

$$\alpha_{i3} = \sin\theta_{i1}\sin\theta_{i2}\cos\theta_{i3},$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\alpha_{ip} = \sin\theta_{i1}\sin\theta_{i2}\sin\theta_{i3}\cdots\sin\theta_{i,p-1},$$

the sets c_{ii}, θ_{ij} which for normal w_{ij} would all be independent with distributions:

$$p(c_{ii}) - \chi^2 \text{ distribution with } p \text{ degrees of freedom,}$$

(12)
$$p(\theta_{ij}) \propto \sin^{p-j-1}\theta_{ij}\,d\theta_{ij},$$

$$(0 \leq \theta_{ij} \leq \pi \text{ for } j = 1, 2, \cdots p-2; \; 0 \leq \theta_{i,p-1} \leq 2\pi),$$

in general retain their independence for given i, but the linkage factor results in an elevation of the χ^2 distributions to n degrees of freedom, and a linkage factor for the θ_{ij} distributions of

(13)
$$|\,\Lambda\,|^{\frac{1}{2}(n-p)} \prod_{i=1}^{p-1} \left\{ \frac{\Gamma[\tfrac{1}{2}(p-i)]\Gamma[\tfrac{1}{2}n]}{\Gamma[\tfrac{1}{2}(n-i)]\Gamma[\tfrac{1}{2}p]} \right\},$$

where

$$\Lambda \equiv \{\alpha_{i1}\alpha_{j1} + \alpha_{i2}\alpha_{j2} + \cdots + \alpha_{ip}\alpha_{jp}\}.$$

We may now, having obtained the distribution of the α_{ij}, note their geometrical interpretation. Let us denote the p components of the dependent variate in n-dimensional sample space by the p vectors ξ_1, ξ_2, $\cdots$, ξ_p. Let the p orthogonal canonical components corresponding to the *sample* canonical correlations r_i be denoted by the p unit vectors x_1, x_2, $\cdots$, x_p. Let the corresponding components for the independent variate be n_i, y_i. The "linkage factor" merely represents the allowance that must be made in the mutual relations of the ξ-vectors for the fact that while they must lie in the p-space of the x-vectors,

they really belong to the original n-space. We may identify the w_{ij} with the coefficients in the equation

(14) $$\xi_i = w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{ip}x_p ,$$

where

$$\xi_i^2 \equiv w_{i1}^2 + w_{i2}^2 + \cdots + w_{ip}^2$$

is a χ^2 with n, and not p, degrees of freedom. If we now suppose for convenience ξ_i to be a unit vector, we have in place of (14)

(15) $$\xi_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + \alpha_{ip}x_p ,$$

with a projection, on the q-space of the y-vectors, of ζ_i , say, where

$$\zeta_i = \alpha_{i1}r_1y_1 + \alpha_{i2}r_2y_2 + \cdots + \alpha_{ip}r_py_p ,$$

and hence, as already noted in the algebraic derivation,

$$R_i^2 = (\xi_i \cdot \zeta_i)^2/\zeta_i^2 = \alpha_{i1}^2r_1^2 + \alpha_{i2}^2r_2^2 + \cdots + \alpha_{ip}^2r_p^2 ,$$

where $(\xi \cdot \zeta)$ denotes a scalar product. The linkage factor (13) indicates that the ξ_i vectors in (15) are not independent in the p-space of the x-vectors, the distribution of their mutual configuration being determined by n-space.

This interpretation enables us to determine the moments of the distribution $p(s_i \mid r_i)$. For if corresponding to (15) we write

(16) $$n_i = \beta_{i1}y_1 + \beta_{i2}y_2 + \cdots + \beta_{iq}y_q ,$$

then

(17) $$s_i = \alpha_{i1}\beta_{i1}r_1 + \alpha_{i2}\beta_{i2}r_2 + \cdots + \alpha_{ip}\beta_{ip}r_p .$$

If we are considering case (a), the relations of the n_i to the y-vectors in q-space will be similar to the relations of the ξ_i to the x-vectors in p-space. In case (b), however, the n_i , which represent the true canonical components of a set of q fixed vectors, must remain strictly orthogonal to each other although their relation to the y-vectors can vary. This means that the relations of the n_i to the y-vectors are determined by a random rotation of a rigid orthogonal set of q vectors in case (b). We may note that if in case (a) we allowed n to tend to infinity, the n_i would also become rigidly orthogonal, so that the solution in case (b) may conveniently be obtained from case (a) by retaining the same distribution of the α_i , and for the β_i letting $n \to \infty$.

Thus in either case the moments of the s_i can be obtained from (17) in terms of the moments of α_{ij} and β_{ij} , two independent sets of coefficients for which the distribution of each set is known. The above comments suffice theoretically to complete the required solution for $(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}$ is a function of α_{ij} and β_{ij} ; the α_{ij} and the corresponding linkage factor can be expressed in terms of $\sin \theta_{ij}$ and $\cos \theta_{ij}$, and similarly for the β_{ij} in terms of, say, $\sin \phi_{ij}$ and $\cos \phi_{ij}$, and integration carried out over the θ_{ij} and ϕ_{ij} . This method is unfortunately

too cumbersome algebraically to be of any practical value except in the case of one non-zero root. This case is considered separately before the general case is discussed further.

5. The case of only one non-zero root. Here we only require $\mu(t)$ and a comparatively simple solution is possible, the linkages within the ξ_i and n_i sets being irrelevant. We have in fact, if ψ is the angle between n_1 and ζ_1, (where ζ_1 was the projection of ξ_1 in the q-space), that ψ is a random angle in the q-space, since the α_{ij} and β_{ij} sets are independent. Hence *in this particular case* we may conveniently write $s_1^2 = R_1^2 \cos^2\psi$, which is just the transformation used to obtain the distribution of the multiple correlation R_1^2. Thus we may replace (10) by (9), where $R_1^2 = \alpha_{11}^2 r_1^2 + \alpha_{12}^2 r_2^2 + \cdots + \alpha_{1p}^2 r_p^2$, and

$$(R_1^2)^t = \sum_{u_1+u_2+\cdots=t} \frac{t! \, (r_1^2)^{u_1} (r_2^2)^{u_2} \cdots}{u_1! \, u_2! \cdots}$$
$$\cdot \cos^{2u_1} \theta_{11} \sin^{2(t-u_1)} \theta_{11} \cos^{2u_2} \theta_{12} \sin^{2(t-u_1-u_2)} \theta_{12} \cdots ,$$

where the expected value of the trigonometric term is evaluated as

$$(18) \qquad \left\{ \frac{\Gamma(u_1 + \frac{1}{2})\Gamma(u_2 + \frac{1}{2})\cdots}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}) \cdots} \right\} \frac{\Gamma(\frac{1}{2}p)}{\Gamma(\frac{1}{2}p + t)} .$$

We have now obtained the distribution, $(\rho_2 = \cdots = \rho_p = 0)$,

$$p(r_i \mid \rho_1 \neq 0) = p(r_i \mid \rho_1 = 0) \sum_{u_1, u_2, \cdots} C(u_1, u_2, \cdots)(r_1^2)^{u_1}(r_2^2)^{u_2} \cdots ,$$

where $p(r_i \mid \rho_i = 0)$ is given by (5); and in case (a)

$$C(u_1, u_2, \cdots) = (1 - \rho_1^2)^{\frac{1}{2}n}(\rho_1^2)^t \left[\frac{\Gamma(\frac{1}{2}n + t)}{\Gamma(\frac{1}{2}n)} \right]^2$$
$$\cdot \frac{\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}{\Gamma(\frac{1}{2}p + t)\Gamma(\frac{1}{2}q + t)} \prod_{j=1}^{p} \left[\frac{\Gamma(u_j + \frac{1}{2})}{\Gamma(\frac{1}{2})u_j!} \right],$$

and in case (b)

$$C(u_1, u_2, \cdots) = e^{-\frac{1}{2}\beta_1^2}(\tfrac{1}{2}\beta_1^2)^t \frac{\Gamma(\frac{1}{2}n + t)\Gamma(\frac{1}{2}p)\Gamma(\frac{1}{2}q)}{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}p + t)\Gamma(\frac{1}{2}q + t)} \prod_{j=1}^{p} \left[\frac{\Gamma(u_j + \frac{1}{2})}{\Gamma(\frac{1}{2})u_j!} \right],$$

where $u_1 + u_2 + \cdots + u_p$ is denoted by t. $\sum_{u_1, u_2, \cdots}$ denotes summation of all u's from 0 to ∞. The solution in either case contains a generalized hypergeometric function. If we denote the general series

$$\sum_{u_1, u_2, \cdots} \left\{ \frac{\Gamma(\alpha_1 + t)\Gamma(\alpha_2 + t)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\Gamma(r_1)\Gamma(r_2)}{\Gamma(r_1 + t)\Gamma(r_2 + t)} \prod_{j=1}^{p} \left[\frac{\Gamma(\beta_j + u_j)x_j^{u_j}}{\Gamma(\beta_j)u_j!} \right] \right\},$$

$$\sum_{u_1, u_2, \cdots} \left\{ \frac{\Gamma(\alpha + t)}{\Gamma(\alpha)} \frac{\Gamma(r_1)\Gamma(r_2)}{\Gamma(r_1 + t)\Gamma(r_2 + t)} \prod_{j=1}^{p} \left[\frac{\Gamma(\beta_j + u_j)x_j^{u_j}}{\Gamma(\beta_j)u_j!} \right] \right\}$$

by

$$F(\alpha_1, \alpha_2 ; \beta_1, \beta_2, \cdots, \beta_p ; r_1, r_2 ; x_1, x_2, \cdots, x_p),$$

$$F(\alpha; \beta_1, \beta_2, \cdots, \beta_p ; r_1, r_2 ; x_1, x_2, \cdots, x_p)$$

respectively (see [8, p. 300, example 22]), then we have in case (a)

(19)
$$p(r_i \mid \rho_1 \neq 0) = p(r_i \mid \rho_i = 0)(1 - \rho_1^2)^{\frac{1}{2}n}$$
$$\times F(\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}; \tfrac{1}{2}p, \tfrac{1}{2}q; \rho_1^2 r_1^2, \rho_1^2 r_2^2, \cdots, \rho_1^2 r_p^2),$$

and in case (b)

(20)
$$p(r_i \mid \beta_1 \neq 0) = p(r_i \mid \rho_i = 0)e^{-\frac{1}{2}\beta_1^2}$$
$$\times F(\tfrac{1}{2}n; \tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}; \tfrac{1}{2}p, \tfrac{1}{2}q; \tfrac{1}{2}\beta_1^2 r_1^2, \tfrac{1}{2}\beta_1^2 r_2^2, \cdots, \tfrac{1}{2}\beta_1^2 r_p^2).$$

An alternative operational form is obtained by noting that the sum of terms for given $t = u_1 + u_2 + \cdots + u_p$ is generated by means of the coefficient of z^t in

$$\prod_{j=1}^{p} (1 - \rho_1^2 r_j^2 z)^{-\frac{1}{2}},$$

where for definiteness we consider case (a). Hence if we write

$$F(\alpha_1, \alpha_2; r_1, r_2; x) \equiv \sum_t \frac{\Gamma(\alpha_1 + t)\Gamma(\alpha_2 + t)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\Gamma(r_1)\Gamma(r_2)}{\Gamma(r_1 + t)\Gamma(r_2 + t)} x^t,$$

we have

(21)
$$F(\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}, \tfrac{1}{2}, \cdots, \tfrac{1}{2}; \tfrac{1}{2}p, \tfrac{1}{2}q; \rho_1^2 r_1^2, \rho_1^2 r_2^2, \cdots, \rho_1^2 r_p^2)$$
$$= \Theta F(\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}p, \tfrac{1}{2}q; z^{-1}) \prod_{j=1}^{p} (1 - \rho_1^2 r_j^2 z)^{-\frac{1}{2}},$$

where Θ denotes the operation of taking the term independent of z (this might possibly be done by multiplication by z^{-1} and evaluation of a suitable contour integral, but in the use of this formula here the operation Θ has been carried out directly).

It is of some interest to examine a simple case, and, incidentally, to check that

$$\int_{r_i} p(r_i \mid \rho_i) = 1.$$

If we take $p = 2$, $q = 3$, we obtain for $p(r_1^2, r_2^2 \mid \rho_1 = \rho_2 = 0)$ the form

$$\tfrac{1}{4}(n - 2)(n - 3)(n - 4)(1 - r_1^2)^{\frac{1}{2}n-3}(1 - r_2^2)^{\frac{1}{2}n-3}(r_1^2 - r_2^2) \, dr_1^2 \, dr_2^2.$$

Considering the distribution (19) with $p = 2$, $q = 3$, and taking the most elementary case $n = 6$, we obtain on integration of r_2^2 from 0 to r_1^2,

$$p(r_1 \mid \rho_1) = 6(r_1^2)^2 \, dr_1^2 (1 - \rho_1^2)^3 \sum_{u_1, u_2} \left[\frac{\Gamma(3 + t)}{\Gamma(3)} \right]^2$$
$$\cdot \frac{\Gamma(\tfrac{3}{2})}{\Gamma(\tfrac{3}{2} + t)} \frac{\Gamma(\tfrac{1}{2} + u_1)\Gamma(\tfrac{1}{2} + u_2)(\rho_1^2 r_1^2)^t}{\Gamma(\tfrac{1}{2})\Gamma(\tfrac{1}{2})u_1!(u_2 + 2)!t!},$$

where $t = u_1 + u_2$. Now from the identity $(1 - x)^{-\frac{1}{2}}(1 - x)^{\frac{1}{2}} = 1 - x$, the

coefficient of x^{t+2}, $(t \geq 0)$, is zero in the expansion of the left-hand side. This provides the identity, for all $t \geq 0$,

$$\sum_{u_1} \frac{\Gamma(\frac{1}{2} + u_1)\Gamma(\frac{1}{2} + t - u_1)\frac{1}{2} \cdot \frac{3}{2}}{\Gamma(\frac{1}{2})u_1!\,\Gamma(\frac{1}{2})(u_2 + 2)!} = \frac{\Gamma(\frac{1}{2} + t + 1)}{\Gamma(\frac{1}{2})(t + 1)!} - \frac{\Gamma(\frac{1}{2} + t + 2)}{\Gamma(\frac{1}{2})(t + 2)!}$$

$$= \frac{\frac{1}{2}(t + 3)\Gamma(\frac{3}{2} + t)}{\Gamma(\frac{1}{2})\Gamma(t + 3)},$$

or

$$(22) \qquad \sum_{u_1} \frac{\Gamma(\frac{1}{2} + u_1)\Gamma(\frac{1}{2} + t - u_1)}{\Gamma(\frac{1}{2})u_1!\,\Gamma(\frac{1}{2})(u_2 + 2)!} = \frac{t + 3}{3} \frac{\Gamma(\frac{3}{2} + t)}{\Gamma(\frac{3}{2})\Gamma(t + 3)}.$$

Hence

$$(23) \qquad \begin{aligned} p(r_1 \mid \rho_1) &= 6(r_1^2)^2\, dr_1^2(1 - \rho_1^2)^3 \sum_t \frac{\Gamma(3 + t)}{[\Gamma(3)]^2} \cdot \frac{(t + 3)}{3}\, (\rho_1^2 r_1^2)^t \\ &= (r_1^2)^2\, dr_1^2(1 - \rho_1^2)^3 \sum_t \frac{\Gamma(3 + t)(\rho_1^2 r_1^2)^t(t + 3)}{\Gamma(3)t!} \\ &= (1 - \rho_1^2)^3\, dr_1^2\, \partial/\partial r_1^2\{(r_1^2)^3(1 - \rho_1^2 r_1^2)^{-3}\}, \end{aligned}$$

which obviously gives unity on integration of r_1^2 from 0 to 1. In purely algebraic form

$$(24) \qquad p(r_1 \mid \rho_1) = 3(1 - \rho_1^2)^3(r_1^2)^2\, dr_1^2/(1 - \rho_1^2 r_1^2)^4.$$

Alternatively, making use of formulae (21), we have for the same case $p = 2$, $q = 3$, $n = 6$, the distribution

$$(25) \quad 6(r_1^2 - r_2^2)\, dr_1^2\, dr_2^2(1 - \rho_1^2)^3\, \Theta F(3, 3; \tfrac{3}{2}; z^{-1})\, (1 - \rho_1^2 r_1^2 z)^{-\frac{1}{2}}\, (1 - \rho_1^2 r_2^2 z)^{-\frac{1}{2}}.$$

Integrating with respect to r_2^2 from 0 to r_1^2, we obtain

$$(26) \quad 6\, dr_1^2(1 - \rho_1^2)^3 \Theta F(3, 3; \tfrac{3}{2}; z^{-1})\, (1 - \rho_1^2 r_1^2 z)^{-\frac{1}{2}} \left\{ \frac{2r_1^2}{\rho_1^2 z} + \frac{4[(1 - \rho_1^2 r_1^2 z)^{\frac{1}{2}} - 1]}{3(\rho_1^2 z)^2} \right\}.$$

Discarding the term for which the irrational expression $(1 - \rho_1^2 r_1^2 z)^{\frac{1}{2}}$ cancels, and hence leaves no terms independent of z, we obtain the distribution $p(r_1 \mid \rho_1)$ given in (23) or (24) by selection of the appropriate terms. We may further integrate directly the expression above with respect to r_1^2, and after discarding again irrelevant terms we obtain

$$(27) \qquad 6(1 - \rho_1^2)^3 \Theta F(3, 3; \tfrac{3}{2}; z^{-1}) \left\{ -\frac{4}{3(\rho_1^2 z)^2}\, (1 - \rho_1^2 z)^{\frac{1}{2}} \right\},$$

which is readily ascertained to be unity.

6. More than one non-zero root. In the general case the factor multiplying $p(r_i \mid \rho_i = 0)$ is rather remarkable in being symmetrical in both the set r_i and

the set ρ_i. As n increases, the convergence of r_1 to ρ_1, r_2 to ρ_2, etc. when the ρ_i are also arranged in descending order of magnitude must result from the restriction $r_1 \geq r_2 \geq \cdots \geq r_p$. The limiting distribution has been discussed by Hsu [9].

In view of the algebraic difficulty of obtaining $\mu(t_1, t_2, \cdots, t_p)$ by direct integration, an unsymmetric method of obtaining the moments was developed. This is fairly tractable in the case of two non-zero roots. The second set w_{2j} of the original variables is transformed by an orthogonal transformation such that the first new variable of the second set is determined by the correlation between w_{1j} and w_{2j}. We may write, for example,

$$w_{21}' = (w_{11}w_{21} + w_{12}w_{22} + \cdots)/(w_{11}^2 + w_{12}^2 + \cdots + w_{1p}^2)^{\frac{1}{2}},$$

$$(28) \qquad w_{22}' = \cfrac{\left\{\cfrac{-w_{11}(w_{12}^2 + \cdots + w_{1p}^2)w_{21}}{w_{11}^2} + w_{12}w_{22} + \cdots\right\}}{\left\{\cfrac{(w_{11}^2 + \cdots + w_{1p}^2)(w_{12}^2 + \cdots + w_{1p}^2)}{w_{11}^2}\right\}^{\frac{1}{2}}},$$

$$\vdots \qquad w_{23}' = \cfrac{\left\{\cfrac{-w_{12}(w_{13}^2 + \cdots + w_{1p}^2)w_{22}}{w_{12}^2} + w_{13}w_{23} + \cdots\right\}}{\left\{\cfrac{(w_{12}^2 + \cdots + w_{1p}^2)(w_{13}^2 + \cdots + w_{1p}^2)}{w_{12}^2}\right\}^{\frac{1}{2}}},$$

which conversely we can at once express as a relation of the w_{2j}, in terms of the w_{2j}', (since the reciprocal of an orthogonal matrix is simply its transpose). If we write

$$(29) \qquad \begin{aligned} \alpha_{21}' &= w_{21}'/[(w_{21}')^2 + (w_{22}')^2 + \cdots + (w_{2p}')^2]^{\frac{1}{2}}, \\ \alpha_{22}' &= w_{22}'/[(w_{21}')^2 + (w_{22}')^2 + \cdots + (w_{2p}')^2]^{\frac{1}{2}}, \\ &\vdots \end{aligned}$$

and write further

$$a_1 = \cos \theta_{11}, a_2 = \cos \theta_{12}, \cdots, b_1 = \cos \theta_{21}',$$

$$b_2 = \cos \theta_{22}', \cdots, \text{where } \alpha_{21}' = \cos \theta_{21}', \alpha_{22}' = \sin \theta_{21}' \cos \theta_{22}', \cdots$$

we have in particular

$$\alpha_{21} = a_1 b_1 - b_2 \sqrt{(1 - a_1^2)} \sqrt{(1 - b_1^2)},$$

$$(30) \quad \alpha_{22} = a_2 b_1 \sqrt{(1 - a_1^2)} + a_1 a_2 b_2 \sqrt{(1 - b_1^2)}$$

$$- b_3 \sqrt{(1 - a_2^2)} \sqrt{(1 - b_1^2)} \sqrt{(1 - b_2^2)},$$

where the distribution of the a's and b's is proportional to

$$\{(1 - a_1^2)^{\frac{1}{2}(p-3)} da_1\}\{(1 - a_2^2)^{\frac{1}{2}(p-4)} da_2\} \cdots \{(1 - b_1^2)^{\frac{1}{2}(n-3)} db_1\}\{(1 - b_2^2)^{\frac{1}{2}(p-4)} db_2\} \cdots.$$

For the reasons discussed in section 4, it will be noticed that only the distribution of b_1 in the a, b set is affected by the linkage factor. By such methods the expressions

$$\mu(1, 1) \equiv E\{s_1^2 s_2^2 \mid r_i\}, \qquad \mu(2, 1) \equiv E\{s_1^4 s_2^2 \mid r_i\}$$

were fairly readily obtained. If we introduce the notation

$$S_k \equiv \sum_{i=1}^{p} (r_i^2)^k, \qquad S_{kl} \equiv \sum_{i \neq j} (r_i^2)^k (r_j^2)^l, \qquad \text{etc.,}$$

and also symbols for the products of the α and β moments, viz.

$$\binom{2}{2} \equiv E\{\alpha_{11}^2 \alpha_{12}^2\} \, E\{\beta_{11}^2 \beta_{12}^2\}, \qquad \binom{2 \quad \cdot}{\cdot \quad 2} \equiv E\{\alpha_{11}^2 \alpha_{22}^2\} \, E\{\beta_{11}^2 \beta_{22}^2\},$$

etc., we may list the moments $\mu(t_1, t_2, \cdots, t_p)$ as in Appendix I, which gives all moments up to the fourth order in terms of the α and β moments (the numerical coefficients arise from the numbers of ways of forming the two-way partitions). "Half-factors" corresponding to the α moments are listed in Appendix II against their appropriate symbol, the corresponding factors coming from the β moments being obtained in case (a) by writing q for p and in case (b) by writing also[5] $n \to \infty$. Thus in case (a)

$$(32) \qquad \begin{aligned} \mu(1, 1) &= \left[\frac{n + 2}{np(p + 2)}\right]\left[\frac{n + 2}{nq(q + 2)}\right] S_2 \\ &\quad + \left\{\left[\frac{np + n - 2}{np(p + 2)(p - 1)}\right]\left[\frac{nq + n - 2}{nq(q + 2)(q - 1)}\right] \right. \\ &\quad \left. + 2\left[\frac{-(n - p)}{np(p + 2)(p - 1)}\right]\left[\frac{-(n - q)}{nq(q + 2)(q - 1)}\right]\right\} 2S_{11}, \end{aligned}$$

and in case (b)

$$(33) \qquad \begin{aligned} \mu(1, 1) &= \left[\frac{n + 2}{np(p + 2)}\right]\left[\frac{1}{q(q + 2)}\right] S_2 \\ &\quad + \left\{\frac{np + n - 2)(q + 1) + 2(n - p)}{np(p + 2)(p - 1)q(q + 2)(q - 1)}\right\} 2S_{11}. \end{aligned}$$

By means of the transformation (28) it is possible to develop the moments $\mu(t_1, t_2)$ in the case of two non-zero roots, though in obtaining the results quoted in Appendix II, where the formula for $\mu(3, 1)$ and $\mu(2, 2)$ are included, it was found convenient to supplement this method with the devices mentioned in the

---

[5] It should be remembered that we have assumed $p \leq q$. If $p > q$, we interchange the dependent and independent vector variates, and hence must interchange p and q in these moment formulae, $p(\leq q)$ now corresponding to the independent variate.

next section. In the case of more than two non-zero roots, it is theoretically possible to carry out a further transformation on the w_{3j} variates, but with the "partial" variates $w_{1j \cdot 2} \equiv w_{1j} - b_{12} w_{2j}$, where

$$b_{12} = (w_{11} w_{21} + w_{12} w_{22} + \cdots) / (w_{11}^2 + w_{12}^2 + \cdots),$$

as coefficients. This enables us to express w_{3j} in terms of new variables, of which the first is related to the partial correlation of w_{3j} with w_{1j} for given w_{2j}, i.e. to the second correlation factor which depends on the "linkage"; and so on. This method is, however, again too cumbersome to be of much use, and a more rapid method of evaluating $\mu(t_1, t_2, \cdots, t_p)$ in general is desirable. This problem has not been entirely solved to the author's satisfaction in this paper, although in the concluding section are mentioned devices which have been found useful, and which enabled the terms for the remaining third-order moment $\mu(1, 1, 1)$ to be completed and added to Appendix II.

7. Relations among the α-moments. Equation (15) defining the α's, the ξ_i being random vectors in the p-space of the x-vectors except for their mutual configuration being determined by the properties of n-space, may be used to provide relations among the α-moments. Thus in addition to the identities

(34) $$\alpha_{i1}^2 + \alpha_{i2}^2 + \cdots + \alpha_{ip}^2 = 1, \qquad (i = 1, 2, \cdots, p),$$

the correlation of any ξ_i with a fixed vector in the p-space, e.g. with x_1 or with $(x_1 + x_2)/\sqrt{2}$, is a random correlation in p-space, whereas the correlation of any ξ_i with any other ξ_i is a random correlation in n-space. The use of these facts is best illustrated by an example and equations sufficient to determine the six α-moments required for $\mu(1, 1, 1)$ will be derived.

For convenience, denote the required mean values of

$$\alpha_{11}^2 \alpha_{21}^2 \alpha_{31}^2 , \quad \alpha_{11}^2 \alpha_{21}^2 \alpha_{32}^2 , \quad \alpha_{11}^2 \alpha_{22}^2 \alpha_{33}^2 , \quad \alpha_{11} \alpha_{12} \alpha_{21} \alpha_{22} \alpha_{31}^2 , \quad \alpha_{11} \alpha_{12} \alpha_{21} \alpha_{22} \alpha_{33}^2 , \quad \alpha_{11} \alpha_{12} \alpha_{22} \alpha_{23} \alpha_{31} \alpha_{33}$$

by A, B, C, D, E, F respectively. Multiply the second-order quantities $\alpha_{11}^2 \alpha_{21}^2$, $\alpha_{11}^2 \alpha_{22}^2$, $\alpha_{11} \alpha_{12} \alpha_{21} \alpha_{22}$ by expression (34) for $i = 3$; since this expression is identically unity, the consequent mean values are unaltered. This gives the three relations

(35)
$$A + (p - 1)B = (n + 2)/\{np(p + 2)\},$$
$$A + 3(p - 1)B + (p - 1)(p - 2)C = 1/p,$$
$$A + (p - 1)B + 2(p - 1)(p - 2)D$$
$$+ (p - 1)(p - 2)(p - 3)E = 1/(np).$$

The moment A is the mean of the triple product of the squared scalar products of ξ_1, ξ_2 and ξ_3 with x_1. The same value must be realized with any other fixed vector in the p-space, e.g. with either $(x_1 + x_2)/\sqrt{2}$ or with $(x_1 + x_2 + \cdots + x_p)/\sqrt{p}$. This gives two relations

(36)
$$A - B - 4D = 0$$
$$(p + 1)A - 3B - 12D - (p - 2)(C + 6E + 8F) = 0.$$

A final linearly independent relation is obtained from the mean triple product of $(\xi_1 \cdot \xi_2)$, $(\xi_1 \cdot \xi_3)$, $(\xi_2 \cdot \xi_3)$, which depends solely on the internal configuration of ξ_1, ξ_2 and ξ_3, and is easily shown (e.g. choose ξ_1 to coincide with one of the original axes of the n-space) to be $1/n^2$. This gives

$$(37) \qquad pA + 3p(p - 1)D + p(p - 1)(p - 2)F = 1/n^2.$$

The equations contained in (35), (36) and (37) determine A, B, C, D, E and F.

Similar equations could evidently be constructed for the higher-order moments, e.g. for the terms required for $\mu(2, 1, 1)$ or $\mu(1, 1, 1, 1)$, but the numbers of such terms increase rapidly. From Appendix I it will be seen that there are 24 distinct terms in $\mu(2, 1, 1)$ and 16 in $\mu(1, 1, 1, 1)$.

Appendix I.

$$\mu(1, 1) = S_2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 2S_{11} \left\{ \begin{pmatrix} 2 & \cdot \\ \cdot & 2 \end{pmatrix} + 2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

$$\mu(2, 1) = S_3 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + S_{21} \left\{ \begin{pmatrix} 4 & \cdot \\ \cdot & 2 \end{pmatrix} + 6 \begin{pmatrix} 2 & 2 \\ 2 & \cdot \end{pmatrix} + 8 \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

$$+ 6S_{111} \left\{ 3 \begin{pmatrix} 2 & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 12 \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & \cdot \end{pmatrix} \right\}$$

$$\mu(3, 1) = S_4 \begin{pmatrix} 6 \\ 2 \end{pmatrix} + S_{31} \left\{ \begin{pmatrix} 6 & \cdot \\ \cdot & 2 \end{pmatrix} + 15 \begin{pmatrix} 4 & 2 \\ 2 & \cdot \end{pmatrix} + 12 \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

$$+ 2 S_{22} \left\{ 15 \begin{pmatrix} 4 & 2 \\ \cdot & 2 \end{pmatrix} + 20 \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} \right\}$$

$$+ 2S_{211} \left\{ 15 \begin{pmatrix} 4 & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 45 \begin{pmatrix} 2 & 2 & 2 \\ 2 & \cdot & \cdot \end{pmatrix} + 120 \begin{pmatrix} 3 & 1 & 2 \\ 1 & 1 & \cdot \end{pmatrix} + 30 \begin{pmatrix} 1 & 1 & 4 \\ 1 & 1 & \cdot \end{pmatrix} \right\}$$

$$+ 24S_{1111} \left\{ 15 \begin{pmatrix} 2 & 2 & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} + 90 \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & \cdot & \cdot \end{pmatrix} \right\}$$

$$\mu(2, 2) = S_4 \begin{pmatrix} 4 \\ 4 \end{pmatrix} + S_{31} \left\{ 12 \begin{pmatrix} 4 & \cdot \\ 2 & 2 \end{pmatrix} + 16 \begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} \right\}$$

$$+ 2S_{22} \left\{ \begin{pmatrix} 4 & \cdot \\ \cdot & 4 \end{pmatrix} + 18 \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} + 16 \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \right\}$$

$$+ 2S_{211} \left\{ 6 \begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 2 & 2 \end{pmatrix} + 36 \begin{pmatrix} 2 & 2 & \cdot \\ 2 & \cdot & 2 \end{pmatrix} + 72 \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \end{pmatrix} + 96 \begin{pmatrix} 3 & 1 & \cdot \\ 1 & 1 & 2 \end{pmatrix} \right\}$$

$$+ 24S_{1111} \left\{ 9 \begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & 2 \end{pmatrix} + 72 \begin{pmatrix} 1 & 1 & 2 & \cdot \\ 1 & 1 & \cdot & 2 \end{pmatrix} + 24 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \right\}$$

$$\mu(1,1,1) = S_3 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + S_{31} \left\{ 3 \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} + 12 \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & \cdot \end{pmatrix} \right\}$$

$$+ 6 S_{111} \left\{ \begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 6 \begin{pmatrix} 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 8 \begin{pmatrix} 1 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{pmatrix} \right\}$$

$$\mu(2,1,1) = S_4 \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix} + S_{31} \left\{ 2 \begin{pmatrix} 4 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} + 6 \begin{pmatrix} 2 & 2 \\ 2 & \cdot \\ 2 & \cdot \end{pmatrix} \right.$$

$$+ 16 \begin{pmatrix} 3 & 1 \\ 1 & 1 \\ 2 & \cdot \end{pmatrix} + 4 \begin{pmatrix} 4 & \cdot \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \Bigg\} + 2 S_{22} \left\{ \begin{pmatrix} 4 & \cdot \\ \cdot & 2 \\ \cdot & 2 \end{pmatrix} + 6 \begin{pmatrix} 2 & 2 \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} \right.$$

$$+ 16 \begin{pmatrix} 3 & 1 \\ 1 & 1 \\ \cdot & 2 \end{pmatrix} + 12 \begin{pmatrix} 2 & 2 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \Bigg\} + 2 S_{211} \left\{ \begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 12 \begin{pmatrix} 2 & 2 & \cdot \\ 2 & \cdot & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} \right.$$

$$+ 3 \begin{pmatrix} \cdot & 2 & 2 \\ 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} + 16 \begin{pmatrix} 3 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 2 \begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 1 & 1 \\ \cdot & 1 & 1 \end{pmatrix} + 24 \begin{pmatrix} 2 & \cdot & 2 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \end{pmatrix}$$

$$+ 48 \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} + 48 \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} + 24 \begin{pmatrix} 2 & 1 & 1 \\ \cdot & 1 & 1 \\ 2 & \cdot & \cdot \end{pmatrix} + 32 \begin{pmatrix} 3 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{pmatrix} \Bigg\}$$

$$+ 24 S_{1111} \left\{ 3 \begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} + 6 \begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} + 24 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{pmatrix} \right.$$

$$+ 24 \begin{pmatrix} 2 & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} + 48 \begin{pmatrix} 1 & 1 & \cdot & 2 \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \end{pmatrix} \Bigg\}$$

$$\mu(1,1,1,1) = S_4 \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} + S_{31} \left\{ 4 \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} + 24 \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & \cdot \\ 2 & \cdot \end{pmatrix} \right\}$$

$$+ 2 S_{22} \left\{ 3 \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \\ \cdot & 2 \end{pmatrix} + 24 \begin{pmatrix} 2 & \cdot \\ \cdot & 2 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} + 8 \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

$$+ 2 S_{211} \left\{ 6 \begin{pmatrix} 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 48 \begin{pmatrix} 2 & \cdot & \cdot \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} + 48 \begin{pmatrix} 1 & \cdot & 1 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} \right.$$

.

$$+ 96 \begin{pmatrix} 1 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \\ 2 & \cdot & \cdot \end{pmatrix} + 12 \begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 1 & 1 \\ \cdot & 1 & 1 \\ 2 & \cdot & \cdot \end{pmatrix} \Bigg\}$$

$$+ 24 S_{1111} \left\{ \begin{pmatrix} 2 & \cdot & \cdot & \cdot \\ \cdot & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} + 12 \begin{pmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} + 32 \begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} \right.$$

$$\left. + 12 \begin{pmatrix} 2 & \cdot & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} + 48 \begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & 1 & 1 \\ 1 & \cdot & \cdot & 1 \end{pmatrix} \right\}.$$

Appendix II.

$$\binom{2}{2} \frac{n+2}{np(p+2)}, \quad \binom{2 \quad \cdot}{\cdot \quad 2} \frac{np+n-2}{np(p+2)(p-1)}, \quad \binom{1 \quad 1}{1 \quad 1} \frac{-(n-p)}{np(p+2)(p-1)},$$

$$\binom{4}{2} \frac{3(n+4)}{np(n+2)(p+4)}, \quad \binom{4 \quad \cdot}{\cdot \quad 2} \frac{3(np+3n-4)}{np(p+2)(p+4)(p-1)},$$

$$\binom{2 \quad 2}{2 \quad \cdot} \frac{np+n+2p-4}{np(p+2)(p+4)(p-1)}, \quad \binom{2 \quad 2 \quad \cdot}{\cdot \quad \cdot \quad 2} \frac{np+3n-4}{np(p+2)(p+4)(p-1)},$$

$$\binom{3 \quad 1}{1 \quad 1} \frac{-3(n-p)}{np(p+2)(p+4)(p-1)}, \quad \binom{1 \quad 1 \quad 2}{1 \quad 1 \quad \cdot} \frac{-(n-p)}{np(p+2)(p+4)(p-1)},$$

$$\binom{6}{2} \frac{15(n+6)}{np(p+2)(p+4)(p+6)}, \quad \binom{6 \quad \cdot}{\cdot \quad 2} \frac{15(np+5n-6)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\binom{4 \quad 2}{2 \quad \cdot} \frac{3(np+n+4p-6)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\binom{4 \quad 2}{\cdot \quad 2} \frac{3(np+3n+2p-6)}{np(p+2)(p+4)(p+6)(p-1)}.$$

$$\binom{4 \quad 2 \quad \cdot}{\cdot \quad \cdot \quad 2} \frac{3(np+5n-6)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\binom{2 \quad 2 \quad 2}{2 \quad \cdot \quad \cdot} \frac{np+3n+2p-6}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\binom{2 \quad 2 \quad 2 \quad \cdot}{\cdot \quad \cdot \quad \cdot \quad 2} \frac{np+5n-6}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\binom{5 \quad 1}{1 \quad 1} \frac{-15(n-p)}{np(p+2)p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} \frac{-9(n-p)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 3 & 1 & 2 \\ 1 & 1 & \cdot \end{pmatrix} \frac{-3(n-p)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & 1 & \cdot \end{pmatrix} \frac{-3(n-p)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & \cdot & \cdot \end{pmatrix} \frac{-(n-p)}{np(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 4 \\ 4 \end{pmatrix} \frac{9(n+4)(n+6)}{n(n+2)p(p+2)(p+4)(p+6)},$$

$$\begin{pmatrix} 4 & \cdot \\ \cdot & 4 \end{pmatrix} \frac{9\{n^2(p+3)(p+5)+2n(p+1)(p+3)-8(2p+3)\}}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 2 & 2 \\ 4 & \cdot \end{pmatrix} \frac{3\{n^2(p+3)+6n(p+1)+8(p-3)\}}{n(n+2)p(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \frac{n^2(p^2+4p+15)+6n(p+1)(p-3)+4(5p^2+2p-6)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 2 & 2 \end{pmatrix} \frac{3\{n^2(p+3)(p+5)+2n(p+1)(p+3)-8(2p+3)\}}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 2 & 2 & \cdot \\ 2 & \cdot & 2 \end{pmatrix} \frac{n^2(p+3)^2+2n(p+1)(2p+3)+4(p^2-4p-6)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & 2 \end{pmatrix} \frac{n^2(p+3)(p+5)+2n(p+1)(p+3)-8(2p+3)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} \frac{-9(n-p)(n+4)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)},$$

$$\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \frac{-9(n-p)(np+3n+2p)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \end{pmatrix} \frac{-(n-p)(np-3n+8p+12)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 3 & 1 & \cdot \\ 1 & 1 & 2 \end{pmatrix} \frac{-3(n-p)(np+3n+2p)}{n(n+2)p(p-2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 1 & 1 & 2 & \cdot \\ 1 & 1 & \cdot & 2 \end{pmatrix} \frac{-(n-p)(np+3n+2p)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \frac{3(n-p)(n-p-2)}{n(n+2)p(p+2)(p+4)(p+6)(p-1)(p+1)},$$

$$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \frac{(n+2)(n+4)}{n^2 p(p+2)(p+4)}, \qquad \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} \frac{(n+2)(np+3n-4)}{n^2 p(p+2)(p+4)(p-1)},$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} \frac{n^2(p^2+3p-2)-6n(p+2)+16}{n^2 p(p+2)(p+4)(p-1)(p-2)},$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & \cdot \end{pmatrix} \frac{-(n-p)(n+2)}{n^2 p(p+2)(p+4)(p-1)},$$

$$\begin{pmatrix} 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} \frac{-(n-p)(np+2n-4)}{n^2 p(p+2)(p+4)(p-1)(p-2)},$$

$$\begin{pmatrix} 1 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{pmatrix} \frac{(n-p)(2n-p)}{n^2 p(p+2)(p+4)(p-1)(p-2)}.$$

REFERENCES

[1] H. HOTELLING, "Relations between two sets of variates", *Biometrika*, Vol. 28 (1936), pp. 321–377.

[2] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations", *Ann. Eugen.*, Vol. 9 (1939), pp. 238–249.

[3] P. L. HSU, "On the distribution of roots of certain determinantal equations", *Ann. Eugen.*, Vol. 9 (1939), pp. 250–258.

[4] S .N. ROY, "*P*-statistics or some generalizations in analysis of variance appropriate to multivariate problems", *Sankhyā*, Vol. 4 (1939), pp. 381–396.

[5] S. N. ROY, "Analysis of variance for multivariate normal populations. The sampling distribution of the requisite *p*-statistics on the null and non-null hypothesis", *Sankhyā*, Vol. 6 (1942), pp. 35–50.

[6] M. S. BARTLETT, "The vector representation of a sample", *Proc. Camb. Phil. Soc.*, Vol. 30 (1934), pp. 327–360.

[7] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient", *Proc. Roy. Soc.*, Vol. A 121 (1928), pp. 654–673.

[8] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, Cambridge Univ. Press, 4th ed., 1935.

[9] P. L. HSU, "On the limiting distribution of the canonical correlations", *Biometrika*, Vol. 32 (1941), pp. 38–45.

ON THE THEORY OF MARKOFF CHAINS

By Elliott W. Montroll

University of Pittsburgh

1. Summary. Although there exists voluminous literature on the theory of probability of independent events, and powerful techniques have been developed for the analysis of most of the interesting problems in this field, the theory of probability of dependent events has been rather neglected. The first detailed investigations in this subject were published by A. Markoff [1]. S. Bernstein [2] has extended the fundamental limit theorems to chains of dependent events. The most extensive exposition of this field has been made by M. Fréchet [3].

In the present paper we shall develop methods of averaging functions over chains of dependent variables and find the probability distribution of these functions. It will be shown that for certain types of chains these averages and distribution functions can be expressed in terms of the characteristic values and vectors of a certain operator equation. Many of the methods discussed here have been applied to problems in statistical mechanics [4, 5, 6, 7, 8]. The most important application has been made by L. Onsager [8] who proved rigorously (on the basis of a simplified model) that Boltzmann's energy distribution in a solid with cooperative elements leads to a phase transition. The first explicit application of linear operator theory (through matrices and integral equations) to probability chains has apparently been made by Hostinsky [9].

2. Introductory Remarks. Suppose there exists a chain of events each of which might lead to one of ν possible results, and which are correlated in such a manner that the probability of n successive events leading to a chain of results

$$\alpha_1 , \alpha_2 , \cdots , \alpha_n$$

is proportional to

$$P_n(\alpha_1 , \alpha_2 , \cdots , \alpha_n).$$

The probability of a given function $F(\alpha_1 , \alpha_2 , \cdots , \alpha_n)$ having a value corresponding to the sequence of α's would be proportional to

$$F(\alpha_1 , \alpha_2 , \cdots , \alpha_n) \, P_n(\alpha_1 , \cdots , \alpha_n)$$

and its average value over all configurations of the chain would be

(1) $$\bar{F} = F_1/F_0 = \sum_{\{\alpha_j\}} F(\alpha_1, \alpha_2 , \cdots , \alpha_n) P_n(\alpha_1, \alpha_2 , \cdots , \alpha_n) / \sum_{\{\alpha_j\}} P_n(\alpha_1 , \cdots , \alpha_n)$$

where

(1a) $$F_m = \sum_{\{\alpha_j\}} [F(\alpha_1 , \alpha_2 , \alpha_3 , \cdots , \alpha_n)]^m \, P_n(\alpha_1 , \cdots , \alpha_n)$$

18

and the summation extends over all values of

$$\{\alpha_j\} = (\alpha_1, \alpha_2, \cdots, \alpha_n).$$

The probability of a result α_1 of the first event leading to a result α_n of the nth event is

$$(2) \qquad P_n(\alpha_1, \alpha_n) = (1/F_0) \sum_{\alpha_2, \cdots, \alpha_{n-1}} P_n(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

In order to find the probability of a given function $F(\alpha_1, \cdots, \alpha_n)$ having a value between ξ and $\xi + h$ it is useful to know the moments and Thiele semi-invariants of $F(\alpha_1, \cdots, \alpha_n)$. Both of these functions of F can be calculated from

$$(3) \qquad Z_n(x) = \sum_{\{\alpha_j\}} P_n(\alpha_1, \cdots, \alpha_n) \exp\{xF(\alpha_1, \alpha_2, \cdots, \alpha_n)\}.$$

Obviously

$$(4) \qquad F_m = \lim_{x \to 0} \partial^m Z_n(x)/\partial x^m.$$

It is known [10] that the mth Thiele semi-invariant is given by

$$(5) \qquad \Lambda_m = \lim_{x \to 0} \partial^m \log Z_n(x)/\partial x^m.$$

In the notation of Cramér $Z_n(i\omega)/Z_n(0) = f(\omega)$, the characteristic function of F.

If $G(z)$ is defined so that $G(\xi + h) - G(\xi)$ is the probability that the function $F(\alpha_1, \cdots, \alpha_n)$ has a value between $\xi \le F(\alpha_1, \cdots, \alpha_n) < \xi + h$, then it is well known that [5] if $G(z)$ is continuous at $x = \xi$ and $x = \xi + h$

$$(6) \qquad G(\xi + h) - G(\xi) = \frac{1}{2\pi i} \lim_{T \to \infty} \int_{-T}^{T} \frac{(1 - e^{-i\omega k})e^{-i\omega\xi}}{\omega} \exp[\log f(\omega)]\, d\omega$$

where

$$(6a) \qquad \log f(\omega) = \sum_{m=1}^{\infty} \frac{\Lambda_m(i\omega)^m}{m!} = \sum_{m=1}^{k} \Lambda_m(i\omega)^m/m! + o(\omega^k).$$

When the derivative of $G(\xi)$ with respect to ξ exists, the probability of

$$F(\alpha_1, \cdots, \alpha_n)$$

having a value between ξ and $\xi + d\xi$ is

$$(6b) \quad \varphi(\xi)\, d\xi = (\partial G/\partial \xi)\, d\xi = \frac{d\xi}{2\pi} \lim_{T \to \infty} \int_{-T}^{T} \exp\left\{\sum_{m=1}^{\infty} \Lambda_m(i\omega)^m/m!\right\} e^{-i\omega\xi}\, d\omega.$$

From (4)

$$(7) \qquad \sum_{m=1}^{\infty} \Lambda_m(i\omega)^m/m! = -\log Z_n(0) + \lim_{x \to 0} e^{-i\omega\partial/\partial x} \log Z_n(x).$$

Since, for a constant c independent of x,

$$e^{c\partial/\partial x} f(x) = f(x + c)$$

we have

$$(8) \qquad \sum_{m=1}^{\infty} \Lambda_m (i\omega)^m / m! = \log \{Z_n(i\omega)/Z_n(0)\},$$

and from (6)

$$(9) \qquad G(\xi + h) - G(\xi) = \frac{1}{2\pi i} \lim_{T \to \infty} \int_{-T}^{T} \frac{e^{-i\omega\xi}(1 - e^{-i\omega h})Z_n(i\omega) \, d\omega}{\omega Z_n(0)} .$$

Equations (3), (4), (5) and (9) indicate that much information concerning a chain of correlated events can be obtained from a knowledge of $Z_n(x)$. We shall now introduce procedures for the determination of $Z_n(x)$ for several general forms of $P(\alpha_1, \cdots, \alpha_n)$.

When α is a continuous variable, the results of this section and those to follow are easily generalized by replacing the summations operations over all values of the α's by integrals, and by replacing the matrix equations of the next section by integral equations.

3. Simple Chains, $P_n(\alpha_1, \cdots, \alpha_n) = \prod_{j=1}^{n-1} p(\alpha_j, \alpha_{j+1})$.

a. General theory. By a simple chain we shall mean a sequence of events, each of which leads to one of ν possible results and which occur in such a manner that if the result of the kth event is α_k, the probability of the $(k + 1)$st one yielding a result α_{k+1} to proportional to $p(\alpha_k, \alpha_{k+1})$. This implies that the probability of the occurrence of the sequence of results

$$\alpha_1, \alpha_2, \cdots, \alpha_n$$

is

$$(10) \qquad \prod_{i=1}^{n-1} p(\alpha_i, \alpha_{i+1}) \Big/ \sum_{\{\alpha_j\}} \prod_{i=1}^{n-1} p(\alpha_i, \alpha_{i+1}),$$

and the probability of a first result α_1, leading to an nth result α_n is

$$(11) \qquad P_n(\alpha_1, \alpha_n) = \sum_{\alpha_2, \cdots, \alpha_{n-1}} \prod_{j=1}^{n} p(\alpha_j, \alpha_{j+1}) \Big/ \sum_{\{\alpha_j\}} \prod_{j=1}^{n} p(\alpha_j, \alpha_{j+1}).$$

The summations are to be extended over all ν possible values of each α_i indicated on the summation indices. Chains of this type are sometimes called simple Markoff chains after the first author who studied them systematically.

From (1), the average value of a function $F(\alpha_1, \cdots, \alpha_n)$ is

$$(12) \qquad F_1/F_0 = \frac{\sum_{\alpha_1} \cdots \sum_{\alpha_n} F(\alpha_1, \cdots, \alpha_n) \prod_{j=0}^{n-1} p(\alpha_j, \alpha_{j+1})}{\sum_{\alpha_1} \cdots \sum_{\alpha_n} \prod_{j=1}^{n-1} p(\alpha_j, \alpha_{j+1})} .$$

Many chain functions $F(\alpha_1, \cdots, \alpha_n)$ of interest are either additive or multiplicative and of one of the forms

(13a) a) $F_1(\alpha_1, \cdots, \alpha_n) = h(\alpha_1, \alpha_2) + h(\alpha_2, \alpha_3) + \cdots + h(\alpha_{n-1}, \alpha_n)$

(13b) b) $F_2(\alpha_1, \cdots, \alpha_n) = g(\alpha_1, \alpha_2) \, g(\alpha_2, \alpha_3) \cdots g(\alpha_{n-1}, \alpha_n)$.

In case (b) it is convenient to define a new function $h(\alpha_i, \alpha_j)$ by

(14) $\qquad g(\alpha_i, \alpha_j) = \exp[xh(\alpha_i, \alpha_j)]$

and in both cases to consider a function of the form

(15) $$Z_n(x) = \sum_{\{\alpha_i\}} \prod_{j=1}^{n-1} p(\alpha_j, \alpha_{j+1}) \exp[xh(\alpha_j, \alpha_{j+1})],$$

for then the values of F_1 and F_2 averaged over the entire chain are given by

(16a) $$<F_1>_{av.} = \lim_{x \to 0} \partial \log Z_n(x)/\partial x$$

and

(16b) $$<F_2>_{av.} = Z_n(1)/Z_n(0).$$

When n is large, the direct evaluation of (15) may become quite difficult because of the large number of variables involved. As an alternative we shall now introduce a procedure that is based on the observation that $Z_n(x)$ is the sum of the elements of the nth power of the matrix

(17) $$\boldsymbol{P}_x = \begin{pmatrix} p_x(1, 1) & p_x(1, 2) & \cdots & p_x(1, \nu) \\ p_x(2, 1) & p_x(2, 2) & \cdots & p_x(2, \nu) \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ p_x(\nu, 1) & p_x(\nu, 2) & \cdots & p_x(\nu, \nu) \end{pmatrix}$$

where the elements $p_x(\alpha, \beta)$ are defined as

(18) $$p_x(\alpha, \beta) = p(\alpha, \beta) \exp[xh(\alpha, \beta)].$$

α and β range over the same set of values as one of the "result" parameters α_j; and each of the ν possible results is represented by a unique integer of the set $1, 2, \cdots, \nu$. Thus $Z_n(x) =$ sum of elements of $\boldsymbol{P}_x^{n-1}$. To employ this observation to advantage, let us consider the characteristic values and vectors of the matrix $\boldsymbol{P}_x$. It is well known that if the characteristic values are simple the characteristic vectors form a biorthogonal set; that is, if

(19a) $$\Phi_{i,x} = \{\varphi_{i,x}(1), \varphi_{i,x}(2), \cdots, \varphi_{i,x}(\nu)\}, \qquad (i = 1, 2, \cdots, \nu),$$

and

(19b) $$\Psi_{i,x} = \begin{bmatrix} \psi_{i,x}(1) \\ \psi_{i,x}(2) \\ \psi_{i,x}(\nu) \end{bmatrix}$$

satisfy the operator equations

$$(20a) \qquad \Phi_{i,x} \cdot P_x = \lambda_{i,x}\Phi_{i,x}$$

$$(20b) \qquad P_x \cdot \Psi_{i,x} = \lambda_{i,x}\Psi_{i,x}$$

where $\lambda_{i,x}$ is the ith characteristic value of (17), then

$$\Phi_{i,x}\cdot\Psi_{j,x} = \sum_{\alpha=1}^{\nu} \varphi_{i,x}(\alpha)\psi_{j,x}(\alpha) = 0 \quad \text{when} \quad i \neq j.$$

We shall for convenience always assume that the φ's and ψ's are normalized:

$$\Phi_{i,x} \cdot \Psi_{i,x} = 1$$

so that in general:

$$(21) \qquad \Phi_{i,x}\cdot\Psi_{j,x} = \delta_{i,j} = \begin{cases} 0 & \text{when} \quad i \neq j \\ 1 & \text{when} \quad i = j. \end{cases}$$

It is well known from matrix theory that one can expand a matrix element as

$$(22) \qquad p_x(\alpha, \beta) = \sum_{i=1}^{\nu} \lambda_{i,x}\varphi_{i,x}(\beta)\psi_{i,x}(\alpha)$$

and that

$$(23) \qquad \lambda_{i,x} = \Phi_{i,x} \cdot P_x \cdot \Psi_{i,x}.$$

By substituting (22) into the expression for $Z_n(x)$ in terms of P_x^{n-1}, one can show that

$$
\begin{aligned}
Z_n(x) &= \sum_{i=1}^{\nu} \{\lambda_{i,x}\}^{n-1}\left\{\sum_{\beta=1}^{\nu} \varphi_{i,x}(\beta)\right\}\left\{\sum_{\alpha=1}^{\nu} \psi_{i,x}(\alpha)\right\} \\
(24) \qquad & \\
&= \sum_{i=1}^{\nu} \lambda_{i,x}^{n-1}(\Phi_{i,x}\cdot 1)(1\cdot\Psi_{i,x}).
\end{aligned}
$$

Therefore $Z_n(x)$ can be determined from a knowledge of the characteristic vectors and values of the matrix P_x.

If there exists a largest characteristic root $\lambda_{i,x}$ such that

$$(25) \qquad \lambda_{L,x} > |\lambda_{i,x}| \qquad\qquad \text{if } i \neq L,$$

one can obtain some interesting results. Before deriving these, we shall give a sufficient condition (which is satisfied in many chains) for the existance of this inequality. Frobenius [11] has shown that if all the elements of a finite matrix are > 0, then the characteristic value of largest absolute value of the matrix is real, positive, and simple (nondegenerate). Thus, as long as ν is finite and $p_x(\alpha, \beta) > 0$ for all α and β, (25) is valid.

We shall now prove that

$$(25a) \qquad \lim_{n\to\infty}\left\{\frac{Z_n(x)}{\lambda_{L,x}^{n-1}(\Phi_{L,x}\cdot 1)(1\cdot\Psi_{L,x})} - 1\right\} = 0$$

that is,

$$(25b) \qquad Z_n(x) \sim \lambda_{L,x}^{n-1}(\Phi_{L,x} \cdot 1)(1 \cdot \Psi_{L,x}).$$

First let us consider the case in which P_x is a symmetrical matrix. Then $\varphi_{j,x}(\alpha) = \psi_{j,x}(\alpha)$, all the characteristic values are real, and

$$Z_n(x) = \lambda_{L,x}^{n-1}(\Phi_{L,x} \cdot 1)^2 + \sum_{i \neq L} \lambda_{i,x}^{n-1}(\Phi_{L,x} \cdot 1)^2.$$

From Cauchy's inequality and (21)

$$|\Phi_{i,x} \cdot 1|^2 = \left| \sum_{\alpha=1}^{\nu} \varphi_{i,x}(\alpha) \right|^2 \leq \left[\sum_{\alpha=1}^{\nu} \varphi_{i,x}^2(\alpha) \right] \left[\sum_{\alpha=1}^{\nu} 1 \right] = \nu.$$

Therefore,

$$\left| \sum_{i \neq L} \lambda_{i,x}^{n-1}(\varphi_{i,x} \cdot 1)^2 \right| \leq \nu \left| \sum_{i \neq L} \lambda_{i,x}^{n-1} \right| \leq \nu(\nu-1) |\lambda_{s,x}^{n-1}|$$

where $\lambda_{s,x}$ is the characteristic value of P_x second largest in absolute value. This inequality yields

$$(25c) \qquad \left| \frac{Z_n(x)}{\lambda_{L,x}^{n-1}(\Phi_{L,x} \cdot 1)^2} - 1 \right| \leq \frac{\nu(\nu-1)}{(\Phi_{L,x} \cdot 1)^2} \left| \left(\frac{\lambda_{s,x}}{\lambda_{L,x}} \right)^{n-1} \right|$$

and (25a) (since $\lambda_{s,x}/\lambda_{L,x} < 1$) follows. When P_x is not symmetrical, one can easily derive the analogous expression

$$\left| \frac{Z_n(x)}{\lambda_{L,x}^{n-1}(\Phi_{L,x} \cdot 1)(1 \cdot \Psi_{L,x})} - 1 \right| \leq \frac{A(\nu-1)|\lambda_{s,x}/\lambda_{L,x}|^{n-1}}{(\varphi_{L,x} \cdot 1)(1 \cdot \psi_{L,x})}$$

where

$$A = [\max \{|(\Phi_{i,x} \cdot 1)|\}][\max \{|(1 \cdot \Psi_{i,x})|\}]$$

For brevity, when $x = 0$, we write $\lambda_{i,x}$ as λ_i, $\Psi_{i,x}$ as Ψ_i, and $\Phi_{i,x}$ as Φ_i. By summing (10) over all α's except α_1, α_k and α_n we obtain the probability of an intermediate event leading to a result α_k if the results of the first and last events are known to have been α_1 and α_n. With the aid of (21) and (22) it is easy to show that this probability is exactly:

$$(26) \qquad \frac{\displaystyle\sum_{i,j=1}^{\nu} \lambda_j^{n-k} \lambda_i^{k-1} \psi_i(\alpha_1)\varphi_i(\alpha_k)\psi_j(\alpha_k)\varphi_j(\alpha_n)}{\displaystyle\sum_{i=1}^{\nu} \lambda_i^{n-1} \sum_{\alpha_1,\alpha_n} \psi_i(\alpha_1)\varphi_i(\alpha_n)}.$$

When n is very large, and when we have simultaneously $n \gg k \gg 1$, we can rewrite this equation to include λ_L, and neglect all terms containing other i's and j's. This leads to the results

    a) If the number of events, n, in a simple chain is very large, the probability $P_n(\alpha_k)$ of a kth event far removed from the first and the last, yielding a result α_k when α_1, and α_n are unspecified is

$$(27) \qquad P_n(\alpha_k) \sim \psi_L(\alpha_k)\,\varphi_L(\alpha_k) \,/\, (\Phi_L \cdot 1)(1 \cdot \Psi_L).$$

b) When $k = n$, the probability of the result $\alpha_1 \cdot$ of the first event leading to the result α_n of the nth event is

$$(28a) \qquad P_n(\alpha_1, \alpha_n) = \frac{\sum_{s=1}^{\nu} \lambda_s^{n-1} \psi_s(\alpha_1)\varphi_s(\alpha_n)}{\sum_{s=1}^{\nu} \lambda_s^{n-1} \sum_{\alpha_1,\alpha_n} \psi_s(\alpha_1)\varphi_s(\alpha_n)} .$$

So, as $n \to \infty$

$$(28b) \qquad P_n(\alpha_1, \alpha_n) \sim \frac{\psi_L(\alpha_1)\phi_L(\alpha_n)}{(\Phi_L \cdot 1)(1 \cdot \Psi_L)} .$$

c) When there exists no knowledge concerning the result of the first event, the probability of the nth event yielding the result α_n is

$$(29) \qquad P_n(\alpha_n) = \sum_{\alpha_1} P_n(\alpha_1, \alpha_n) \sim \Phi_L(\alpha_n)/(1 \cdot \Phi_L).$$

In chains of sufficient length for (25) to be valid, the probability of

$$F(\alpha_1, \cdots, \alpha_n)$$

having a value between ξ and $\xi + h$ has an especially simple asymptotic form. From (6) this probability is (if for a given n we let $T = an^{\frac{1}{2}}$)

$$(30) \qquad G(\xi + h) - G(\xi) = \frac{1}{2\pi i} \lim_{a \to \infty} \int_{-an^{1/2}}^{an^{1/2}} \left(\frac{d\omega}{\omega}\right) e^{-i\omega(\xi - \Lambda_1)}$$
$$(1 - e^{-i\omega h}) \exp\left\{ -\tfrac{1}{2}\omega^2 \Lambda_2 - \frac{i\Lambda_3 \omega^3}{3!} + \cdots \right\}$$

and from (25) and (5)

$$(31) \qquad \Lambda_m \sim n \lim_{x \to 0} \partial^m \log \lambda_{L,x}/\partial x^m = nL_m$$

if

$$(32) \qquad L_m \equiv \lim_{x \to 0} \partial^m \log \lambda_{L,x}/\partial x^m.$$

Letting $y = \omega n^{\frac{1}{2}}$, (30) becomes

$$(33) \qquad G(\xi + h) - G(\xi) \sim \frac{1}{2\pi i} \lim_{a \to \infty} \int_{-a}^{a} \frac{dy}{y}$$
$$(e^{-iy\mu_1} - e^{-iy\mu_2})e^{-\frac{1}{2}y^2 L_2} \left\{ 1 - \frac{L_3 y^3 i}{6n^{\frac{1}{2}}} + \cdots \right\}$$

where

$$(34a) \qquad \mu_1 = (\xi - \Lambda_1)/n^{\frac{1}{2}}$$
$$\mu_2 = (\xi + h - \Lambda_1)/n^{\frac{1}{2}}$$

$$(34b) \qquad \Lambda_1 = \text{average value of } F(\alpha_1, \cdots, \alpha_n) = \bar{F}.$$

Integrating (33)

$$(35) \qquad G(\xi + h) - G(\xi) \sim \frac{1}{(2\pi L^2)^{\frac{1}{2}}} \int_{\mu_1}^{\mu_2} e^{-\mu^2/2L_2}[1 + O(1/n)] \, d\mu.$$

As $n \rightarrow \infty$ and $h \rightarrow 0$

$$(35a) \qquad G(\xi + h) - G(\xi) \sim \frac{h}{(2\pi n L_2)^{\frac{1}{2}}} \exp (-\tfrac{1}{2})[\xi - \bar{F}]/nL_2),$$

and the probability that $\xi \leq F < \xi + h$ becomes Gaussian.

b. *Examples of a simple chain.* As an example of a simple Markoff chain let us consider an event which can lead to either of two possible results, say "-1" or "1". Further, let us suppose that the probability of a given result being followed by an identical one is p and by one of another type is $(1 - p)$; that is,

$$p(-1, -1) = p(1, \quad 1) = p$$
$$p(-1, \quad 1) = p(1, -1) = 1 - p.$$

This chain would be encountered in an analysis of a sequence of tosses of a coin with a "memory" so that the probability of two successive tosses showing the same face of the coin would be p and that of showing opposite faces $(1 - p)$.

A question one might ask concerning such a chain is—What is the probability of the occurrence of a given number of transitions from one kind of result to another? In the chain of results

$$-1, -1, -1, 1, 1, -1, 1, -1, -1, -1$$

there would be four transitions, one corresponding to each -1 followed by a 1 and to each 1 followed by a -1. The function giving the number of transitions in a sequence of n events is

$$(36) \qquad F(\alpha_1, \cdots, \alpha_n) \doteq \sum_{i=1}^{n-1} h(\alpha_i, \alpha_{i+1})$$

where

$$h(-1, -1) = h(1, \quad 1) = 0$$
$$h(-1, \quad 1) = h(1, -1) = 1.$$

Even though the α's are dependent, in this special case, $h(\alpha_i, \alpha_{i+1})$ and $h(\alpha_{i+1}, \alpha_{i+2})$ are independent so that (40) could have been obtained on this basis.

To apply the methods described in the beginning of this section we must find the characteristic values and vectors of the matrix

$$(37) \qquad P_x = \begin{pmatrix} p & (1 - p)e^x \\ (1 - p)e^x & p \end{pmatrix}$$

(the configuration index α has the value either -1 or 1 in this case instead of

"1" and "2" as given in (17)).   The characteristic values are the roots of the equation

$$\begin{vmatrix} p - \lambda & (1 - p)e^z \\ (1 - p)e^z & p - \lambda \end{vmatrix} = 0$$

that is,

(38) $\lambda_{1,z} = p + (1 - p)e^z$

$$| \lambda_{2,z} | = | p - (1 - p)e^z | < \lambda_{1,z}$$

and the characteristic vectors are

$$\psi_{1,z} = 2^{-\frac{1}{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \psi_{2,z} = 2^{-\frac{1}{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The ψ and φ vectors have the same components in this case because of the symmetry of the P_z matrix.   Clearly

$$\lambda_L = \lambda_1 = \lambda_{1,0} = 1; \qquad \lambda_2 = \lambda_{2,0} = 2p - 1$$

$$\psi_1(\alpha) = 2^{-\frac{1}{2}} \quad \text{and} \quad \psi_2(\alpha) = -\alpha \cdot 2^{-\frac{1}{2}}.$$

From (26) we see that if the result of the first event in the chain is α_1, and that of the nth event is α_n, the probability of the kth event yielding the result α_k is

$$\frac{[(2p - 1)^{k-1} \alpha_1 \alpha_k + 1][1 + (2p - 1)^{n-k} \alpha_k \alpha_n]}{2[1 + (2p - 1)^{n-1} \alpha_1 \alpha_n]}.$$

As k, n_1 and $(n - k)$ simultaneously get very large, $P_n(\alpha_k) \sim \frac{1}{2}$, independently of α_k.

The probability of an initial result α_1 leading to a final result α_n is (from 28a)

$$P_n(\alpha_1, \alpha_n) = (\tfrac{1}{2}) \{1 + (2p - 1)^{n-1} \alpha_1 \alpha_n\}$$

so that

$$P_n(\ 1, 1) = P_n(-1, -1) = (\tfrac{1}{4}) \{1 + (2p - 1)^{n-1}\}$$

$$P_n(-1, 1) = P_n(\ 1, -1) = (\tfrac{1}{4}) \{1 - (2p - 1)^{n-1}\}.$$

Now, to answer our original question regarding the probability distribution of the transition function (36)

(39) $F(\alpha_1, \cdots, \alpha_n) = \sum_{i=1}^{n-1} h(\alpha_i, \alpha_{i+1}),$

we use the expression for $Z_n(x)$ determined from (24)

(39) $Z_n(x) = 2[p + (1 - p)e^z]^{n-1}$

From (9) the probability of there being between ξ and $\xi + h$ transitions in a sequence of $n + 1$ events is

$$G(\xi + h) - G(\xi) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{i\omega\xi}(1 - e^{i\omega\xi})\{p + (1 - p)e^{i\omega}\}^n \, d\omega/\omega$$

(40)

$$= \frac{1}{2\pi i} \int_{-\infty}^{\infty} (e^{-i\omega\xi} - e^{-i\omega(\xi-h)}) \sum_{k=0}^{n} \frac{n!(1 - p)^k p^{n-k}}{(n - k)!k!}.$$

Letting $x = \omega h/2$ and rearranging

$$G(\xi + h) - G(\xi) = \frac{1}{\pi} \sum_{k=0}^{n} \frac{n!(1 - p)^k p^{n-k}}{(n - k)!k!} D\left(1 + \frac{2}{h}(\xi + h)\right),$$

where $D(\lambda)$ is the Dirichlet integral

$$D(\lambda) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin x \cos \lambda x}{x} \, dx = \frac{1}{2} \quad \begin{array}{ll} 0 & \text{if } |\lambda| > 1 \\ \frac{1}{2} & |\lambda| = 1 \\ 1 & |\lambda| < 1. \end{array}$$

We therefore have, when $[\xi + h] \leq n$

(41)
$$G(\xi + h) - G(\xi) = \sum_{k=[\xi+1]}^{[\xi+h]} \frac{n!(1 - p)^k p^{n-k}}{(n - k)!k!}.$$

Here $[x]$ denotes the greatest integer not exceeding x. The sum is zero if $[\xi + h] < [\xi + 1]$. When $[\xi + h] > n$

(42)
$$G(\xi + h) - G(\xi) = \sum_{k=[\xi+1]}^{n} \frac{n!(1 - p)^k p^{n-k}}{k!(n - k)!}.$$

When n is large it is difficult to get a clear picture of the function $G(\xi)$ from (41) and (42), so we shall develop asymptotic results for large n by using (6) instead of (9).

By employing (5), we see that (this section will be developed on the basis of $n + 1$ trials instead of n)

$$\Lambda_1 = \bar{F} = n(1 - p)$$

$$\Lambda_2 = np(1 - p)$$

$$\Lambda_3 = np(1 - p)(2p - 1) \text{ etc.}$$

Therefore, from (6)

$$\Delta G = G(\xi + h) - G(\xi) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{-i\omega(\xi-\Lambda_1)}(1 - e^{-i\omega h})}{\omega}$$

$$\exp\left[-\tfrac{1}{2}np(1 - p)\omega^2 - inp(1 - p)(2p - 1)\omega^3/6 - \cdots\right] d\omega.$$

Letting $u = \omega n^{\frac{1}{2}}$, we have

$$\Delta G = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{du}{u} \left[e^{-iu(\xi - \Lambda_1)/n^{\frac{1}{2}}} - e^{-iu(\xi + h - \Lambda_1)/n^{\frac{1}{2}}} \right]$$

$$\left[1 - \frac{ip(1-p)(2p-1)u^3}{6n^{\frac{1}{2}}} + O\left(\frac{u^4}{n}\right) \right] e^{-\frac{1}{2}u^2 p(1-p)}$$

$$\dot{=} \frac{1}{2\pi} \int_{\mu_1}^{\mu_2} d\lambda \int_{-\infty}^{\infty} e^{-iu\lambda} \left[1 - \frac{ip(1-p)(2p-1)u^3}{6n^{\frac{1}{2}}} + O\left(\frac{u^4}{n}\right) \right] e^{-\frac{1}{2}u^2 p(1-p)} \, du$$

where

$$\mu_1 = (\xi + h - \Lambda_1)/n^{\frac{1}{2}}$$

$$\mu_2 = (\xi - \Lambda_1)/n^{\frac{1}{2}}.$$

Since

$$\int_{-\infty}^{\infty} e^{-au^2} e^{-i\lambda u} \, du = (\pi/a)^{\frac{1}{2}} \exp\left(-\lambda^2/4a\right)$$

$$i \int_{-\infty}^{\infty} u^3 e^{-au^2} e^{-i\lambda u} \, du = \frac{3\lambda \pi^{\frac{1}{2}}}{4a^{5/2}} \left(1 - \frac{\lambda^2}{6a}\right) e^{-\lambda^2/4a},$$

we have for large n

(43a)
$$\Delta G \sim \frac{1}{[2\pi p(1-p)]^{\frac{1}{2}}} \int_{-\mu_1}^{\mu_2} e^{-\lambda^2/2p(1-p)}$$

$$\left\{ 1 - \frac{(2p-1)\lambda}{2p(1-p)n^{\frac{1}{2}}} \left(1 - \frac{\lambda^2}{3p(1-p)}\right) + O\left(\frac{1}{n}\right) \right\} d\lambda.$$

As $n \to \infty$ and $h \to 0$, this becomes

(43b)
$$G(\xi + h) - G(\xi) \sim \frac{h \exp\left\{-[\xi - \bar{F}]^2/2p(1-p)n\right\}}{[2\pi np(1-p)]^{\frac{1}{2}}}$$

$$\left\{ 1 - \frac{(2p-1)(\xi - \bar{F})}{2p(1-p)n} + O\left(\frac{1}{n^2}\right) \right\}.$$

A similar problem which occurs in statistics of high polymers can be stated abstractly as follows. Suppose there exists a sequence of events each of which leads to a translation of length a of a point either to the right or to the left, and that the probability of a translation continuing in the same direction as its predecessor is p while that of changing its direction is $(1 - p)$. After n translations what is the probability of a point being displaced a distance ξ from its origin.

If "-1" represents a translation to the left and "$+1$" a translation to the right,

$$p(-1, -1) = p(1, \quad 1) = p$$

$$p(-1, \quad 1) = p(1, -1) = (1 - p)$$

The function giving the distance of the point from its origin after n displacements is (when $\alpha = \pm 1$)

$$F(\alpha_1, \cdots, \alpha_n) = a \sum_{j=1}^{n} \alpha_j = \tfrac{1}{2}a\alpha_1 + h(\alpha_1, \alpha_2) + \cdots + h(\alpha_{n-1}, \alpha_n) + \tfrac{1}{2}a\alpha_n$$

where

$$h(1, \quad 1) = a, \quad h(-1, -1) = -a$$
$$h(1, -1) = h(-1, 1) = 0.$$

Neglecting the terms $a\alpha_1/2$ and $a\alpha_n/2$ in $F(\alpha_1, \cdots, \alpha_n)$, one can answer questions concerning this problem by evaluating $Z_n(x)$ as defined by (15). In this case P_x has the form

$$P_x = \begin{pmatrix} pe^{ax} & 1 - p \\ 1 - p & pe^{-ax} \end{pmatrix}.$$

Its characteristic roots are

$$\lambda_{1,x} = p \cosh ax + [p^2 \cosh^2 ax + (1 - 2p)]^{\frac{1}{2}} = \lambda_{L,x}$$
$$|\lambda_{2,x}| = |p \cosh ax - [p^2 \cosh^2 ax + (1 - 2p)]^{\frac{1}{2}}| < \lambda_1, x.$$

and its characteristic vectors:

$$\psi_{1,x} = [(p - 1)^2 + (pe^{ax} - \lambda_1)^2]^{-\frac{1}{2}} \begin{pmatrix} p - 1 \\ pe^{ax} - \lambda_1 \end{pmatrix}$$

$$\psi_{2,x} = [(p - 1)^2 + (pe^{ax} - \lambda_2)^2]^{-\frac{1}{2}} \begin{pmatrix} p - 1 \\ pe^{ax} - \lambda_2 \end{pmatrix}.$$

Since

$$\bar{F} = \Lambda_1 = \lim_{x \to 0} \partial \log Z_n(x)/\partial x,$$

one can show in the present problem that $\bar{F} = 0$. Therefore, the probability of the translated point being a distance between ξ and $\xi + h$ from the origin after $(n + 1)$ translations, is, as $n \to \infty$ and $h \to 0$

$$F(\xi + h) - F(\xi) \sim h(2\pi n L_2)^{-\frac{1}{2}} e^{-\xi^2/2nL_2}$$

where L_2 is by (32):

$$L_2 = \lim_{x \to 0} \partial^2 \log \lambda_{L,x}/\partial x = a^2 p/(1 - p).$$

Thus,

$$F(\xi + h) - F(\xi) \sim h[a^2 2\pi np/(1 - p)]^{-\frac{1}{2}} e^{-\xi^2(1-p)/2na^2 p}.$$

When $p = 2/3$ this problem is equivalent to the determination of the proba-

bility distribution of the components in an arbitrary direction of the distance between the ends of a linear polymer. In this case

$$F(\xi + h) - F(\xi) \sim h(4a^2\pi n)^{-\frac{1}{2}} \exp{(-\xi^2/4na^2)}$$

a result obtained by Tobolsky [12] after a lengthy and complicated combinatory calculation.

Another type of simple chain is encountered in the determination of the "life span" of a particle which is displaced a unit distance to the right or left per unit time along a straight line until it collides with an absorbing boundary either $-(q + 1)$ or $(p + 1)$ units from the starting point. This problem has been analyzed by M. Kac using the methods discussed in the present paper. We shall generalize his results to include the effect of an attraction of the particle toward one end of the line so that displacements toward that end are more probable than those in the other direction.

Following the notation of Kac [13] we let X_j represent the jth displacement, m_j its length, and $\delta(m)$ the probability of a given displacement having the length m. Then,

$$\delta(m) = \begin{array}{ll} s & \text{if } m = 1 \\ 1 - s & \text{if } m = -1 \\ 0 & \text{otherwise.} \end{array}$$

If N represents the life span of a particle, the probability of its exceeding n is

$$\text{Prob } \{N > n\} = \text{Prob } \{-q \leq X_1 \leq p, -q \leq X_1 + X_2 \leq p, \cdots,$$

$$-q \leq X_1 + X_2 + \cdots + X_n \leq p\} = \Sigma \, \delta(m_1)\delta(m_2) \cdots \delta(m_n)$$

where the summation extends over all integers $m_1, m_2, \cdots, m_n$ such that

$$-q \leq m_1 \leq p, -q \leq m_1 + m_2 \leq p, \cdots, -q \leq m_1 + m_2 + \cdots + m_n \leq p.$$

Defining the new set of variables

$$\alpha_j = q + m_1 + m_2 + \cdots + m_j \qquad (j = 1, 2, \cdots n)$$

we see that

$$\text{Prob } \{N > n\} = \sum_{\alpha_1, \cdots, \alpha_n = 0}^{p+q} \delta(\alpha_1 - q)\delta(\alpha_2 - \alpha_1) \cdots \delta(\alpha_n - \alpha_{n-1}).$$

As before, if we introduce the P matrix (of $p + q + 1$ rows and columns)

$$P = (\delta(\alpha - \beta)) = \begin{Bmatrix} 0 & 1-s & 0 & 0 & \cdots \\ s & 0 & 1-s & 0 & \cdots \\ 0 & s & 0 & 1-s & \cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{Bmatrix}$$

we obtain after applying the equivalent of (22)

$$\text{Prob } \{N > n\} = \sum_{j=1}^{p+q+1} \lambda_j^n \varphi_j(q) \sum_{\alpha_n=0}^{p+q} \psi_j(\alpha_n).$$

Where λ_j is the jth characteristic value of P, and ψ_j and φ_j are its associated characteristic vectors as defined by (19) and (20) (here the range of α starts from 0 instead of 1 as in (17) and (19)).

It is easy to show that the characteristic values of P are

$$\lambda_j = 2[s(1 - s)]^{\frac{1}{2}} \cos \zeta_j, \quad (j = 1, 2, \cdots, p + q + 1)$$

where

$$\zeta_j = \pi j/(p + q + 2)$$

and that the components of the characteristic vectors are

$$\psi_j(\alpha) = [2/(p + q + 2)]^{\frac{1}{2}} [s/(1 - s)]^{\frac{1}{2}\alpha} \sin (\alpha + 1)\zeta_j, \quad (\alpha = 0, 1, \cdots, p + q)$$

and

$$\varphi_j(\alpha) = [2/(p + q + 2)]^{\frac{1}{2}} [(1 - s)/s]^{\frac{1}{2}\alpha} \sin (\alpha + 1)\zeta_j .$$

Since

$$\sum_{\alpha_n=0}^{p+q} \psi_j(\alpha_n) = \frac{\sqrt{2} (1 - s)}{\sqrt{p + q + 2}} \frac{\{1 - 1(-1)^j[s/1 - s]^{\frac{1}{2}(p+q+2)}\} \sin \zeta_j}{1 - 2[s(r - s)]^{\frac{1}{2}} \cos \zeta_j}$$

we finally have

$$\text{Prob } \{N > n\} = \frac{(1 - s)^{\frac{1}{2}(n+q+2)} 2^{n+1} s^{\frac{1}{2}(n-q)}}{p + q + 2}$$
$$\sum_{j=1}^{p+q+1} \frac{\{1 - (-1)^j(s/1 - s)^{\frac{1}{2}(p+q+2)}\} \cos^n \zeta_j \sin \zeta_j \sin (q + 1)\zeta_j}{1 - 2\sqrt{s(1 - s)} \cos \zeta_j}$$

When $s = \frac{1}{2}$ this reduces to the result of Kac: (* means summation is only over even j's

$$\text{Prob } \{N > n\} = \frac{2}{p + q + 2} \sum_{j=1}^{p+q+1} {}^* \cos^n \zeta_j \sin (q + 1)\zeta_j \cot \frac{1}{2}\zeta_j .$$

4. Simple Chains with Restrictions.

Often when studying chains of dependent events, certain functions averaged over the entire chains are known to be restricted between definite limits. That is, there might exist k functions $g_j(\alpha_1, \alpha_2, \cdots, \alpha_n)$ such that

$$(44) \qquad -\Delta G_j < G_j - g_j(\alpha_1, \cdots, \alpha_n) < \Delta G_j, \quad (j = 1, 2, \cdots k),$$

where the G_j's and ΔG_j's are preassigned constants. To calculate averages of other functions (1) is no longer valid, for it is an unrestricted sum over all sets

of α's, including those incompatible with (44). All unrestricted sums in this formula (and other similar ones) must be replaced by sums over only those sets of α's compatible with (44). Since it is sometimes more difficult to evaluate restricted sums than unrestricted ones, we shall apply an idea of Markoff [1] to the reduction of the former to the latter type.

Let us seek an explicit expression for a function $P_n^*(\alpha_1, \alpha_2, \cdots, \alpha_n)$ which has the property:

$$P_n^*(\alpha_1, \cdots, \alpha_n) = P_n(\alpha_1, \cdots, \alpha_n) \qquad \begin{array}{l} \text{when } \alpha\text{'s are chosen} \\ \text{so that (44) is satis-} \\ \text{fied of all } j. \end{array}$$
$$0 \qquad\qquad\qquad \text{otherwise.}$$

Since the Dirichlet integrals

$$\delta_j = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin (\rho_j \Delta G_j)}{\rho_j} \exp (i\rho_j \gamma_j) \, d\rho_j$$

have the property

$$\delta_j = 1 \text{ when } -\Delta G_j < \gamma_j < \Delta G_j$$

$$0 \text{ otherwise,}$$

$$P_n^*(\alpha_1, \cdots, \alpha_n) = \delta_1 \delta_2 \cdots \delta_k P_n(\alpha_1, \cdots, \alpha_n)$$

has the required character provided

$$\gamma_j = G_j - g_j(\alpha_1, \cdots, \alpha_n).$$

The average value of a function $F(\alpha_1, \cdots, \alpha_n)$ can be written in terms of the unrestricted sum

$$\bar{F} = \sum_{\{\alpha_s\}} F(\alpha_1, \cdots, \alpha_n) P_n^*(\alpha_1, \cdots, \alpha_n) / \sum_{\{\alpha_s\}} P_n^*(\alpha_1, \cdots, \alpha_n),$$

where the summation extends over the complete set of $\{\alpha_s\}$'s

$$\{\alpha_s\} = (\alpha_1, \alpha_2, \cdots, \alpha_n).$$

As in the case of chains without auxiliary restrictions, a useful function is

$$Z_n(x) = \sum_{\{\alpha_s\}} P_n^*(\alpha_1, \cdots, \alpha_n) \exp \{xF(\alpha_1, \cdots, \alpha_n)\}$$

(45)

$$= \frac{1}{\pi^k} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S_n(x, \rho_1, \cdots, \rho_k) \prod_{m=1}^{k} \left\{ \frac{\sin (\rho_m \Delta G_m)}{\rho_m} e^{i\rho_m G_m} \, d\rho_m \right\}$$

where

$$S_n(x, \rho_1, \cdots, \rho_k) = \sum_{\{\alpha_s\}} P_n(\alpha_1, \cdots, \alpha_n)$$

$$\exp \left\{ xF(\alpha_1, \cdots, \alpha_n) - i \sum_{j=1}^{k} \rho_j g_j(\alpha_1, \cdots, \alpha_n) \right\}.$$

When $F(\alpha_1, \cdots, \alpha_n)$ and $\{g_j(\alpha_1, \cdots, \alpha_n)\}$ are all additive or multiplicative functions of the form (13a) and (13b), say

$$F(\alpha_1, \cdots, \alpha_n) = \sum_{k=1}^{n-1} h(\alpha_k, \alpha_{k+1})$$

$$g_j(\alpha_1, \cdots, \alpha_n) = \sum_{k=1}^{n-1} g_j(\alpha_k, \alpha_{k+1})$$

and the probability chain is a simple one, $Z_n(x)$ reduces to a simple form.
Suppose

$$P_n(\alpha_1, \cdots, \alpha_n) = \sum_{j=1}^{n-1} p(\alpha_j, \alpha_{j+1})$$

then following the derivation of (24), we have

(46)
$$S_n(x, \rho_1, \cdots, \rho_k) = \sum_{l=1}^{\nu} \{\lambda_{l,x,\rho}\}^{n-1} (\Phi_{l,x,\rho} \cdot 1)(1 \cdot \Psi_{l,x,\rho})$$

where $\lambda_{l,x,\rho}$, $\Phi_{l,x,\rho}$ and $\Psi_{l,x,\rho}$ are characteristic values and vectors of the matrix

$$P_{x,\rho} = \begin{pmatrix} p_{x,\rho}(1, 1) \cdots p_{x,\rho}(1, \nu) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p_{x,\rho}(\nu, 1) \cdots p_{x,\rho}(\nu, \nu) \end{pmatrix}$$

and

$$p_{x,\rho}(\alpha, \beta) = p(\alpha, \beta) \exp \{xh(\alpha, \beta) - i \sum_{j} \rho_j g_j(\alpha, \beta)\}.$$

Substitution of (46) into (45) allows one to calculate $Z_n(x)$.

5. More Complicated Chains. In a chain of N events in which the result of each event depends on those of its n predecessors ($n << N$), the calculation of $Z_n(x)$ proceeds in essentially the same manner as in the case of a simple chain. Let the N events be divided into N/n sets of "grand events" of n simple events each (for simplicity we assume N is divisible by n, this can easily be avoided). Thus, if each simple event could lead to any one of ν possible results, a grand event could lead to any one of ν^n possible results and a complicated chain becomes a simple chain of grand events with the result of each grand event depending on the preceeding grand event. Quantitative calculations thus proceed formally in the same manner as in a simple chain.

6. Continuous Case. In this section we generalize, by studying an example, to the case in which each event in a simple chain may lead to any one of a continuum of results. The example is a problem arising in statistical mechanics of molecular chains.

Consider a linear chain of n identical molecules whose centers of mass remain at a set of fixed regularly spaced positions, but which may rotate about their

centers of mass in a plane. Suppose, that the potential energy of interaction between neighboring pairs of molecules is a function of the angles a specified axis of the molecules makes with the line connection the centers of mass of the molecules; that is, the potential energy of interaction between pairs of adjacent molecules can be written as $V(\theta_j, \theta_{j+1})$. Assuming that forces are sufficiently short ranged for interaction between more distant neighbors can be neglected, Boltzmann's theorem states that the probability of the axis of the first molecule making an angle between θ_1 and $\theta_1 + a\theta_1$ with the line of centers of the chain, the second between θ_2 and $\theta_2 + \alpha\theta_2$ and the nth between θ_n and $\theta_n + d\theta_n$ is proportional to

$$\exp\left[-kT\left\{V(\theta_1, \theta_2) + V(\theta_2, \theta_3) + \cdots + V(\theta_{n-1}, \theta_n)\right\}\right] d\theta_1 \cdots d\theta_n$$

where k is Boltzmann's constant and T is the absolute temperature. The contribution of the interaction to the thermodynamic properties of the chain can be derived from the partition function

$$
(47) \quad
\begin{aligned}
Z_n &= \int_0^{2\pi} \int_0^{2\pi} \cdots \int_0^{2\pi} \\
&\quad \exp\left\{-\frac{1}{kT}\left[V(\theta_1, \theta_2) + \cdots + V(\theta_{n-1}, \theta_n)\right]\right\} d\theta_1 \cdots d\theta_n.
\end{aligned}
$$

For example, the internal energy is

$$\bar{E} = \partial \log Z_n / \partial(-1/kT)$$

and the specific heat is $c = \partial E/\partial T$.

It is to be noted that Z_n is exactly the integral of the iterated kernel of the integral equation

$$(48) \qquad \lambda\psi(\theta_1) = \int_0^{2\pi} \psi(\theta_2) \exp\left\{-\frac{1}{kT} V(\theta_1, \theta_2)\right\} d\theta_2.$$

If $V(\theta_1, \theta_2)$ is symmetrical in θ_1 and θ_2, this linear homogeneous integral equation has a set of orthonormal characteristic functions $\{\psi_j(\theta)\}$ such that

$$(49) \qquad \int_0^{2\pi} \psi_j(\theta)\psi_k(\theta)\, d\theta = \delta_{jk}.$$

To each of these characteristic functions there corresponds a characteristic value λ_j. Now it is well known that the kernel of (48) can be expanded as a series in its characteristic functions

$$\exp\left\{-\frac{1}{kT} V(\theta_1, \theta_2)\right\} = \sum_j \lambda_j \psi_j(\theta_1)\psi_j(\theta_2).$$

Introduction of this expression into (47) and applying the orthogonality conditions (49) one obtains

$$(47a) \qquad Z_n = \sum_j \lambda_j^{n-1}\left\{\int_0^{2\pi} \psi_j(\theta)\, d\theta\right\}^2.$$

Probably the most interesting example of a molecular chain of the type described above is a chain of magnetic dipoles which are restricted to rotate only in a plane.   In that case

$$V(\theta_j, \theta_{j+1}) = \frac{\mu^2}{r^3}[\cos(\theta_j - \theta_{j+1}) - 3\cos\theta_j\cos\theta_{j+1}].$$

Where μ is the magnetic moment of each dipole and r is the distance between a pair of adjacent centers of mass.   This potential function leads to the integral equation

$$\lambda\psi(\theta_1) = \int_0^{2\pi} \psi(\theta_2)\exp\left\{-\frac{\mu^2}{r^3 kT}[\cos(\theta_1 - \theta_2) - 3\cos\theta_1\cos\theta_2]\right\} d\theta_2.$$

Since this equation is rather complicated to solve, we shall devote the rest of the section to a potential function of less physical interest, but which leads to a less formidable integral equation.

In studying hindered rotation of molecules, one sometimes uses potential functions of the form:

$$V(\theta_j, \theta_{j+1}) = -\beta\cos(\theta_j - \theta_{j+1})$$

where β is a constant.   With this potential function (48) becomes

(50) $$\lambda\psi(\theta_1) = \int_0^{2\pi} \psi(\theta_2)\exp\{J\cos(\theta_1 - \theta_2)\} d\theta_2$$

where

$$J = \beta/kT.$$

The characteristic functions and characteristic values of (50) are easily found with the aid of the Fourier Series for $\exp(J\cos\theta)$:

(51) $$\exp(J\cos\theta) = I_0(J) + 2\sum_{m=1}^{\infty} I_m(J)\cos m\theta$$

where $I_m(J)$ is the mth Bessel function of imaginary argument:

$$I_m(J) = \sum_{m=0}^{\infty} \frac{(\frac{1}{2}J)^{2k+m}}{(m+k)!k!}.$$

From (51)

$$\exp[J\cos(\theta_1 - \theta_2)] = I_0(J) + 2\sum_{m=1}^{\infty} I_m(J)(\cos m\theta_1\cos m\theta_2 + \sin m\theta_1\sin m\theta_2).$$

Substituting this expression into (50) we have

$$\lambda\psi(\theta_1) = \int_0^{2\pi} \psi(\theta_2)\left\{I_0(J) + 2\sum_{m=1}^{\infty} I_m(J)(\cos m\theta_1\cos m\theta_2 + \sin m\theta_1\sin m\theta_2)\right\} d\theta_2.$$

Because of the orthogonality of the trigonometric functions, one can verify by direct substitution that the characteristic functions are

$$\psi_0(\theta) = 1/(2\pi)^{\frac{1}{2}}$$

$$\psi_m^{(1)}(\theta) = \pi^{-\frac{1}{2}} \sin m\theta; \qquad \psi_m^{(2)} = \pi^{-\frac{1}{2}} \cos m\theta, \ (m = 1, 2, \cdots)$$

and the corresponding characteristic values are

$$\lambda_0 = 2\pi I_0(J)$$

$$\lambda_m^{(1)} = \lambda_m^{(2)} = 2\pi I_m(J) \qquad\qquad\qquad m > 0.$$

Introduction of these characteristic functions and values into (47a) we obtain the simple formula for the partition function:

$$Z_n = 2\pi \{2\pi I_0(J)\}^{n-1}.$$

The internal energy of the molecular chain is therefore

$$\bar{E} = \partial \log Z_n / \partial(-1/kT)$$

$$= -\beta(n - 1) \ I_1(J)/I_0(J),$$

and the specific heat is:

$$C = \partial\bar{E}/\partial T = \tfrac{1}{2}k(n - 1)J^2 \left\{ 1 + \frac{I_2(J)}{I_0(J)} - 2\left[\frac{I_1(J)}{I_0(J)}\right]^2 \right\}.$$

REFERENCES

[1] A. MARKOFF, *Wahrscheinlichkeitsrechnung*, Leipzig, 1912.
[2] S. BERNSTEIN, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dependantes," *Math. Ann.*, Vol. 97 (1927), p. 1.
[3] M. FRECHÉT, *Recherces Theoretiques Moderns sur La Theorie des Probabilites*, Vol. 2, Paris (1937).
[4] H. KRAMERS AND G. WANNIER, "Statistics of the two-dimensional ferromagnet: Part I," *Phys. Rev.*, Vol. 60, (1941), p. 252.
[5] E. MONTROLL, "Statistical mechanics of nearest neighbor systems," *Jour. Chem. Phys.*, Vol. 9 (1941), p. 708; Vol. 10 (1942), p. 61.
[6] E. LASSETTRE AND J. HOWE, "Thermodynamic properties of binary solid solutions on the basis of the nearest neighbor approximation," *Jour. Chem. Phys.*, Vol. 9 (1941), p. 747.
[7] J. ASHKIN AND W. E. LAMB, "The propagation of order in crystal lattices," *Phys. Rev.*, Vol. 64 (1943), p. 159.
[8] L. ONSAGER, "Crystal statistics I. A two-dimensional model with an order-disorder transition," *Phys. Rev.*, Vol. 65, (1944), p. 117.
[9] M. HOSTINSKY, *Methodes generales du Calculu des Probabilites*, Paris, 1931.
[10] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Univ. Press, 1937, Chap. 4.
[11] G. FROBENIUS, "Über Matrizen aus positiven Elementen. II.," *Preuss. Acad. Wiss. Sitz.*, (1909), p. 514.
[12] A. TOBOLSKY, POWELL AND H. EYRING, an article in *Chemistry of Large Molecules*, Interscience Publishers, 1943, pp. 156, 182.
[13] M. KAC, "Random walk in the presence of absorbing barriers," *Annals of Math. Stat.* Vol. 14, (1945), p. 62.

ON THE FIRST TWO MOMENTS OF THE MEASURE OF A RANDOM SET

BY L. A. SANTALÓ

Universidad Nacional del Litoral, Argentina

1. Introduction. In a recent paper [3] H. E. Robbins derived general formulas for the moments of the measure of any random set X, and applied the formulas to find the mean and the variance of a random sum of intervals on a line. In subsequent papers, J. Bronowski and J. Neyman [1], using other methods, found the variance when X is a random sum of rectangles in the plane, and H. E. Robbins [4] found the variance when X is a random sum of n-dimensional intervals in n-dimensional euclidean space. In the latter paper Robbins solved also the corresponding problem for circles on the plane.

Using the methods of Robbins, our purpose in the present paper is to solve the following similar problems:

(i) Let R denote the rectangle consisting of all points (x,y) such that $0 \leq x \leq A_1$, $0 \leq y \leq A_2$, and let R' denote the larger rectangle for which $-\delta \leq x \leq A_1 + \delta$, $-\delta \leq y \leq A_2 + \delta$. Let ρ denote a rectangle of fixed dimensions, $a \times b$, but variable position in the plane. The position of ρ will be determined by the coordinates x, y of its center P and the angle φ between the side of length a and the x-axis. We suppose $(a^2 + b^2)^{\frac{1}{2}} \leq \min (A_1, A_2, \delta)$. Let a fixed number N of rectangles ρ be chosen independently with the probability density function for the coordinates (x, y, φ) of each rectangle constant and equal to $\frac{1}{2} \pi R'$ in the three-dimensional interval with base R' and height π and zero outside this interval. In section 3 we evaluate the first two moments of the measure of X, where X denotes the intersection of the set-theoretical sum of the N rectangles ρ with R.

(ii) Let R denote the n-dimensional interval consisting of all points $(x_1, x_2, x_3, \cdots, x_n)$ such that $0 \leq x_i \leq A_i$, $(i = 1, 2, \cdots, n)$, and let R' denote the larger interval for which $-\delta \leq x_i \leq A_i + \delta$. Let a fixed number N of n-dimensional spheres with radii r (such that $2r \leq \min (A_i, 2\delta)$) be chosen independently, with the probability density function for the centre of each n-sphere constant and equal to $1/R'$ in R' and zero outside this interval. Denoting by X the intersection of the set theoretical sum of the N n-spheres with R, we evaluate in section 4 the first two moments of the measure of X. This problem is a generalization to n-dimensional space of the case considered by Robbins for the plane $(n = 2)$ in [4].

2. Preliminary formulas. Let K be an indeformable plane convex figure of variable position in the plane. The position of K may be determined by the coordinates (x, y) of a point P fixed within K and the angle φ which measures the rotation of K about P. We shall call x, y, φ the coordinates of K. The

measure of a set of figures congruent with K is defined as being the integral of the differential form

$$(2.1) \qquad\qquad dK = dxdyd\varphi.$$

It is readily shown that this measure does not depend on the particular point P chosen to determined the position of K[5]. For instance, the measure of the set of figures K, each of which contains in its interior a fixed point Q, has the value $2\,\pi F$, where F denotes the area of K; that is,

$$(2.2) \qquad\qquad \int_{Q \in K} dK = 2\pi F.$$

Let P_1 and P_2 be two fixed points and let l be the distance P_1P_2. The measure of the set of figures congruent with K, each of which contains both points P_1 and P_2 in its interior, will be a function of K and l, say $\mu(K, l)$. If d is the diameter of K, that is, the maximal distance between two points of K, we have $\mu(K, l) = 0$ for $l \geq d$.

Examples. Let K be a rectangle ρ of fixed dimensions $a \times b$, and let us suppose $a \leq b$. The diameter d of ρ is $d = (a^2 + b^2)^{\frac{1}{2}}$. Let $P(x, y)$ be the centre of ρ and φ the angle which forms the side of length b with the segment line P_1P_2 of length l. If we keep first φ constant, then in order that there exist positions of ρ in which it contains the segment line P_1P_2 in its interior it is necessary that

$$a - l \sin \varphi \geq 0, \qquad b - l \cos \varphi \geq 0$$

and in this case the area covered by the centres P in all these positions has the value

$$(a - l \sin \varphi) (b - l \cos \varphi).$$

Integrating over all permissible values of φ, we obtain

$$(2.3) \qquad \mu(\rho, l) = 4 \int_{\text{arc cos}[b/l]_1}^{\text{arc sin}[a/l]_1} (a - l \sin \varphi)(b - l \cos \varphi)\, d\varphi$$

where we define

$$[x]_1 = \begin{array}{l} x \text{ if } x \leq 1 \\ 1 \text{ if } x \geq 1. \end{array}$$

Carrying out the obvious integration in (2.3) we have

$$(2.4) \qquad \mu(\rho, l) = \begin{cases} 2\,\pi ab - 4\,l(a + b) + 2\,l^2 & \text{for } l \leq a \leq b \\ 4(ab \text{ arc sin } (a/l) - \tfrac{1}{2}\,a^2 - bl + b(l^2 - a^2)^{\frac{1}{2}}) & \\ & \text{for } a \leq l \leq b \\ 4(ab \text{ arc sin } (a/l) - \text{arc cos } (b/l) + b(l^2 - a^2)^{\frac{1}{2}} & \\ + a(l^2 - b^2)^{\frac{1}{2}} - \tfrac{1}{2}(a^2 + b^2) - \tfrac{1}{2}\,l^2) & \text{for } a \leq b \leq l. \end{cases}$$

As another example, let R be the rectangle consisting of all points (x, y) such that $0 \leq x \leq A_1$, $0 \leq y \leq A_2$ and let R' be the rectangle consisting of all points (x, y) such that

$$-\delta \leq x \leq A_1 + \delta, \quad -\delta \leq y \leq A_2 + \delta, \quad (a^2 + b^2)^{\frac{1}{2}} \leq \min(A_1, A_2, \delta).$$

Let us consider the set of rectangles ρ whose centers belong to R' and do not contain either P_1 or P_2, P_1 and P_2 being two fixed points which belong to R. Let l be the distance $P_1 P_2$. According to (2.2) and the definition of $\mu(\rho, l)$ the measure of the set of rectangles ρ under consideration is

$$(2.5) \qquad 2\pi R' - 2.2\pi\rho + \mu(\rho, l),$$

where $R' = (A_1 + 2\delta)(A_2 + 2\delta)$ and $\rho = ab$.

Let K be a plane convex figure of fixed position in its plane. Let us suppose K to be translated a distance l in the direction θ, and let $F(Km, l, \theta)$ be the area of the intersection of K with the translated figure. Obviously if d is the diameter of K, $F(K, l, \theta) = 0$ for $l \geq d$. In what follows we shall consider the function

$$(2.6) \qquad \Phi(K, l) = \int_0^{2\pi} F(K, l, \theta)\, d\theta.$$

Example. Let K be a rectangle R of sides A_1, A_2. Let the symbol $[x]$, as in [1], be defined by

$$[x] = \begin{array}{l} x \text{ if } x \geq 0 \\ 0 \text{ if } x \leq 0. \end{array}$$

It is then readily seen that

$$(2.7) \qquad F(R, l, \theta) = [A_1 - l\sin\theta][A_2 - l\cos\theta].$$

For our purpose the case in which $l \leq \min(A_1, A_2)$ is of interest. In this case, carrying out the immediate integrations, we obtain

$$(2.8) \qquad \Phi(R, l) = 2\pi A_1 A_2 - 4l(A_1 + A_2) + 2l^2.$$

Let $S_{n,r}$ be an n-dimensional sphere of radius r. $S_{n,r}$ will denote also the volume of this sphere, that is, as is known, (see [2, p. 109]),

$$(2.9) \qquad S_{n,r} = \frac{(\pi r^2)^{n/2}}{\Gamma\left(\dfrac{n}{2} + 1\right)}.$$

Let us call the measure of a set of spheres $S_{n,r}$ the measure of the set of their centers. That is, if the point $P(x_1, x_2, \cdots, x_n)$ is the center of $S_{n,r}$ the measure of a set of spheres $S_{n,r}$ equals the integral extended over the set, of the differential form

$$(2.10) \qquad dP = dx_1 dx_2 \cdots dx_n.$$

For instance, the measure of the set of spheres $S_{n,r}$, each of which contains a fixed point Q in its interior, has the value

$$(2.11) \qquad \int_{Q \in S_{n,r}} dP = S_{n,r}$$

where $S_{n,r}$ is given by (2.9).

The measure $\mu(S_{n,r}, l)$ of the set of spheres $S_{n,r}$, each of which contains totally in its interior a segment of length $l(l \leq 2r)$, equals the volume of the intersection of two-spheres $S_{n,r}$ whose centers are placed at the end points of the given segment. That is, $\mu(S_{n,r}, l)$ equals twice the volume of the spherical segment of an n-sphere of radius r and semiangle $\alpha = \text{arc } \cos (l/2r)$. We will represent the volume of this spherical segment by $S_{n,r}(\alpha)$ and it may be calculated in the following way: The intersection of the n-sphere with a hyperplane at a distance x from the center is an $(n - 1)$-dimensional sphere of radius $r' = (r^2 - x^2)^{\frac{1}{2}}$. Let $S_{n-1,r'}$ denote the volume of this $(n - 1)$-dimensional sphere (given by the general formula (2.9)). The volume of the spherical segment, whose base has the radius $h = r \cos \alpha$, will be

$$S_{n,r}(\alpha) = \int_h^r S_{n-1,r'} \, dx.$$

Putting $x = r \cos \theta$ and substituting for $S_{n-1,r'}$ the expression given in (2.9), we obtain

$$S_{n,r}(\alpha) = \frac{\pi^{(n-1)/2} r^n}{\Gamma\left(\dfrac{n+1}{2}\right)} \int_0^\alpha \sin^n \theta \, d\theta = r S_{n-1,r} \int_0^\alpha \sin^n \theta \, d\theta.$$

Consequently we can write

$$(2.12) \qquad \mu(S_{n,r}, l) = 2 S_{n,r}(\alpha) = 2r S_{n-1,r} \int_0^\alpha \sin^n \theta \, d\theta,$$

where $S_{n-1,r}$ is the volume of the $(n - 1)$-dimensional sphere of radius r and $\alpha = \text{arc } \cos (l/2r)$.

In (2.12) we may substitute

$$\int_0^\alpha \sin^n \theta \, d\theta = \frac{(n-1)(n-3) \cdots 3.1}{n(n-2) \cdots 4.2} \text{ arc } \cos (l/2r)$$

$$(2.13) \qquad - \frac{l}{2r} \left\{ \frac{1}{n} \left(1 - \frac{l^2}{4r^2}\right)^{(n-1)/2} + \frac{(n-1)}{n(n-2)} \left(1 - \frac{l^2}{4r^2}\right)^{(n-3)/2} \right.$$

$$\left. + \cdots + \frac{(n-1)(n-3) \cdots 3.1}{n(n-2) \cdots 4.2} \left(1 - \frac{l^2}{4r^2}\right)^{\frac{1}{2}} \right\}$$

for n even, and

$$
\begin{aligned}
(2.14) \quad \int_0^\alpha \sin^n \theta \, d\theta \frac{(n-1)(n-3)\cdots 4.2}{n(n-2)\cdots 3} &- \frac{l}{2r}\left\{ \frac{1}{n}\left(1 - \frac{l^2}{4r^2}\right)^{(n-1)/2} \right. \\
&\left. + \frac{n-1}{n(n-2)}\left(1 - \frac{l^2}{4r^2}\right)^{(n-3)/2} + \cdots + \frac{(n-1)(n-3)\cdots 4.2}{n(n-2)\cdots 5.3} \right\}
\end{aligned}
$$

for n odd.

In particular, for $n = 2, 3$ we have

$$
(2.15) \quad \mu(S_{2,r},l) = 4r^2 \int_0^\alpha \sin^2 \theta \, d\theta = 2r^2 \arccos (l/2r) - \frac{1}{2} l(4r^2 - l^2)^{\frac{1}{2}}
$$

$$
(2.16) \quad \mu(S_{3,r},l) = 2\pi r^2 \int_0^\alpha \sin \theta \, d\theta = \frac{4}{3}\pi r^3 - \pi r^2 l + \frac{1}{12}\pi l^3.
$$

We shall now generalize the formula (2.8) to n-space.

A direction in n-space may be given by the corresponding point on the surface of the n-dimensional sphere of unit radius, that is, by the end point of the radius which is parallel to the given direction. The parametric equations of the n-sphere $\sum_1^n \xi_i^2 = 1$ are

$$
\begin{aligned}
\xi_1 &= \cos \varphi_1 \\
\xi_2 &= \sin \varphi_1 \cos \varphi_2 \\
(2.17) \quad \xi_3 &= \sin \varphi_1 \sin \varphi_2 \cos \varphi_3 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\xi_{n-1} &= \sin \varphi_1 \sin \varphi_2 \cdots \sin \varphi_{n-2} \cos \varphi_{n-1} \\
\xi_n &= \sin \varphi_1 \sin \varphi_2 \cdots \sin \varphi_{n-2} \sin \varphi_{n-1},
\end{aligned}
$$

where $0 \le \varphi_i \le \pi$ for $i < n - 1$ and $0 \le \varphi_{n-1} \le 2\pi$. The element of area of this n-sphere has the value (see, [2, p. 109])

$$
(2.18) \quad d\sigma = \sin^{n-2} \varphi_1 \sin^{n-3} \varphi_2 \cdots \sin \varphi_{n-2} \, d\varphi_1 d\varphi_2 \cdots d\varphi_{n-1}.
$$

A direction in n-dimensional space may then be given by the $n - 1$ parameters $\varphi_1, \varphi_2, \cdots, \varphi_{n-1}$.

Given the n-dimensional interval R consisting of all points $(x_1, x_2, x_3, \cdots, x_n)$ such that $0 \le x_i \le A_i$ $(i = 1, 2, 3, \cdots n)$, and suppose that R is translated a distance $l(l \le \min (A_1, A_2, A_3, \cdots, A_n))$ in the direction $(\varphi_1, \varphi_2, \cdots, \varphi_{n-1})$, the intersection of the translated interval with R is a new interval whose volume has the value $\prod_1^n (A_i - x_i)$, where $x_i = l\xi_i$ (ξ_i given by (2.17)).

Our purpose is to evaluate the integral

$$(2.19) \qquad \Phi(R, l) = \int_{E_n} \prod_1^n (A_i - x_i) \, d\sigma$$

extended over the surface E_n of the n-dimensional sphere of radius unity. We shall denote by E_m either the surface of the m-dimensional sphere of radius unity or its area, given, as is known [2, p. 110] by

$$(2.20) \qquad E_m = \frac{2\pi^{m/2}}{\Gamma\left(\dfrac{m}{2}\right)} \, .$$

Because of the symmetry, the coefficients of all the products $A_{i_1} A_{i_2} A_{i_3} \cdots A_{i_{n-k}}$ have the same value

$$\alpha_k = (-1)^k \int_{E_n} x_1 x_2 \cdots x_k \, d\sigma.$$

The integral extended over the whole surface E_n equals 2^n times the integral extended over the portion for which $\xi_i \geq 0$. Hence, taking into account (2.17) and (2.18) we get

$$\alpha_k = (-1)^k 2^k l^k E_{n-k} \int_0^{\pi/2} \cdots \int_0^{\pi/2} \sin^{n+k-3}\varphi_1 \cos \varphi_1 \sin^{n+k-5}\varphi_2 \cos \varphi_2$$

$$(2.21) \qquad\qquad\qquad \cdots \sin^{n-k-1}\varphi_k \cos \varphi_k \, d\varphi_1 \, d\varphi_2 \cdots d\varphi_k$$

$$= (-1)^k \frac{2^k l^k E_{n-k}}{(n + k - 2)(n + k - 4) \cdots (n + k - 2k)}$$

for $k = 1, 2, \cdots, n - 1$. For $k = n$ we find that

$$\alpha_n = (-1)^n 2^n l^n \int_0^{\pi/2} \cdots \int_0^{\pi/2} \sin^{2n-3}\varphi_1 \cos \varphi_1$$

$$(2.22) \qquad\qquad\qquad \cdots \sin \varphi_{n-1} \cos \varphi_{n-1} \, d\varphi_1 \, d\varphi_2 \cdots d\varphi_{n-1}$$

$$= (-1)^n \frac{2^n l^n}{(2n - 2)(2n - 4) \cdots 4.2} \, .$$

Hence, we have the following general formula

$$\Phi(R, l) = A_1 A_2 \cdots A_n E_n + (-1)^n \frac{2^n l^n}{(2n - 2)(2n - 4) \cdots 4.2}$$

$$(2.23) \qquad\qquad + \sum_{k=1}^{n-1} (-1)^k \Big(\sum_{i_1, i_2, \cdots, i_{n-k}} A_{i_1} A_{i_2} \cdots A_{i_{n-k}} \Big)$$

$$\frac{2^k l^k E_{n-k}}{(n + k - 2)(n + k - 4) \cdots (n + k - 2k)} \, .$$

In particular, for $n = 2$ this result coincides with (2.8). For $n = 3$ we have

$$\Phi(\dot{R}, l) = 4\pi A_1 A_2 A_3 - l^3 - 2\pi l(A_1 A_2 + A_1 A_3 + A_2 A_3)$$

$$(2.24)$$
$$+ \tfrac{8}{3} l^2 (A_1 + A_2 + A_3).$$

3. First problem. We can now solve the first problem (i) stated in the introduction. Denoting by the same letters either sets or their measures, we consider, as in [1] and [4], the set Y of points of R that do not belong to X. We have identically:

$$(3.1) \qquad\qquad X + Y = R.$$

The general method of Robbins [3] taking into account (2.2), gives immediately the first moments

$$(3.2) \qquad E(Y) = R\left(1 - \frac{\rho}{R'}\right)^N, \qquad E(X) = R\left\{1 - \left(1 - \frac{\rho}{R'}\right)^N\right\},$$

where $R = A_1 A_2$, $R' = (A_1 + 2\delta)(A_2 + 2\delta)$, $\rho = ab$.

Our remaining problem is that of evaluating the second moment of X. Let x_i, y_i, φ_i $(i = 1, 2, 3, \cdots, N)$ be the coordinates of the N rectangles ρ (section 2) and let us put, as in (2.1), $d\rho_i = dx_i dy_i d\varphi_i$. Let $P(x, y)$ and $P_0(x_0, y_0)$ be two points which belong to R and let us put $dP = dx\,dy$, $dP_0 = dx_0 dy_0$. Let us consider the following multiple integral

$$(3.3) \qquad\qquad J = \int \frac{dP\,dP_0\,d\rho_1\,d\rho_2\cdots d\rho_N}{(2\pi R')^N}$$

extended over the sets of rectangles ρ_i (congruent with ρ) such that x_i, y_i belongs to R', $0 \le \varphi_i \le 2\pi$, and do not contain either P or P_0. That is, the domain of integration of J is defined by

$$(3.4) \qquad \begin{array}{cccc} -\delta \le x_i \le A_1 + \delta, & -\delta \le y_i \le A_2 + \delta, & 0 \le \varphi_i \ge 2\pi, \\[4pt] P \in R, & P_0 \in R, & P \notin \rho_i, & P_0 \notin \rho_i, & (i = 1, 2, \cdots, N). \end{array}$$

In order to calculate J, we can first keep the rectangles ρ_i fixed; the points P and P_0 can then vary independently over the set of points Y. That gives

$$(3.5) \qquad\qquad J = \int\limits_{(x_i, y_i) \in R'} \frac{Y^2\,d\rho_1\,d\rho_2\cdots d\rho_N}{(2\pi R')^N} = E(Y^2).$$

We can now reverse the order of integration, an operation which is obviously justified in this case. Keeping P and P_0 fixed, we can vary each rectangle ρ_i over the set of positions in which it does not contain either P or P_0; letting l denote the distance PP_0, we have, according to (2.5),

$$(3.6) \qquad\qquad J = \int\limits_{P \in R, P_0 \in R} \left(1 - \frac{4\pi\rho - \mu(\rho, l)}{2\pi R'}\right)^N dP\,dP_0.$$

In order to evaluate this integral we divide it into two parts $J = J_1 + J_2$, according as $0 \leq l \leq d$ or $d \leq l \leq D$, where $d = (a^2 + b^2)^{\frac{1}{2}}$ and $D = (A_1^2 + A_2^2)^{\frac{1}{2}}$. In the interval $0 \leq l \leq d$ we introduce the new variables of integration l, θ related to x, y, x_0, y_0 by

$$(3.7) \qquad x_0 = x + l \cos \theta, \qquad y_0 = y + l \sin \theta$$

whence

$$\frac{\partial(x, y, x_0, y_0)}{\partial(x, y, l, \theta)} = l.$$

In terms of the new variables we have

$$J_1 = \int \left(1 - \frac{4\pi p - \mu(p, l)}{(2\pi R')}\right)^N l \, dl \, dP \, d\theta.$$

In this integral the point P can vary over the intersection of R with the figure obtained by translating R a distance l in the direction θ; that is, the integration of dP gives the function $F(R, l, \theta)$ defined in section 2. According to (2.6) we therefore have

$$(3.8) \qquad J_1 = \int_0^d \left(1 - \frac{4\pi\rho - \mu(\rho, l)}{2\pi R'}\right)^N \Phi(R, l)l \, dl,$$

where $\mu(\rho, l)$ is given by (2.4) and $\Phi(R, l)$ by (2.8).

In order to evaluate J_2 we observe that in the interval $d \leq l \leq D$ $\mu(\rho, l) = 0$ and we have

$$J_2 = \left(1 - \frac{2\rho}{R'}\right)^N \int\limits_{d \leq l \leq D} dP \, dP_0 = \left(1 - \frac{2\rho}{R'}\right)^N \left\{\int^{0 \leq l \leq d} dP \, dP_0 - \int\limits_{0 \leq l \leq d} dP \, dP_0\right\}.$$

Further we have

$$(3.9) \qquad \int\limits_{0 \leq l \leq D} dP \, dP_0 = R^2$$

and with the change of variables (3.7) and the formula (2.8) we find that

$$(3.10) \qquad \int\limits_{0 \leq l \leq d} dP \, dP_0 = \int_0^d \Phi(R, l)l \, dl = \pi A_1 A_1 d^2 - \frac{4}{3} (A_1 + A_2) d^3 + \frac{1}{2} d^4.$$

Collecting (3.8), (3.9), (3.10) and taking into account (3.5) we have

$$(3.11) \qquad \begin{aligned} E(Y^2) &= \int_0^d \left(1 - \frac{4\pi\rho - \mu(\rho, l)}{2\pi R'}\right)^N \Phi(R, l)l \, dl \\ &\quad + \left(1 - \frac{2\rho}{R'}\right)^N \{R^2 - \pi A_1 A_2 d^2 + \tfrac{4}{3}(A_1 + A_2) d^3 - \tfrac{1}{2}d_4\}, \end{aligned}$$

where $\rho = ab, R = A_1A_2 , R' = (A_1 + 2\delta)(A_2 + 2\delta), \mu(\rho, l)$ is given by (2.4) and $\Phi(R, l)$ by (2.8).

For the variance of X and of Y, we have by (3.1) and (3.2)

$$\sigma_X^2 = E(X^2) - E^2(X) = E(Y^2) - E^2(Y)$$

$$= \int_0^d \left(1 - \frac{4\pi\rho - \mu(\rho, l)}{2\pi R'}\right)^N \Phi(R, l)l\,dl + \left(1 - \frac{2\rho}{R'}\right)^N$$

$$\cdot \{R^2 - \pi A_1 A_2 d^2 + \tfrac{4}{3}(A_1 A_2)d^3 - \tfrac{1}{2}d^4\} - R^2\left(1 - \frac{\rho}{R'}\right)^{2N}$$

which completes the solution of our first problem stated in the introduction.

4. Second problem. In order to solve the second problem (ii) stated in the introduction we will follow the same method of the preceding section.

Let X be the intersection of the set theoretical sum of the N n-dimensional spheres $S_{n,r}$ of radius r with the n-interval R. Let us call Y the set of those points of R that do not belong to X, that is,

$$(4.1) \qquad\qquad X + Y = R.$$

The general method of Robbins gives immediately

$$(4.2) \qquad E(Y) = R\left(1 - \frac{S_{n,r}}{R'}\right)^N, \qquad E(X) = R\left\{1 - \left(1 - \frac{S_{n,r}}{R'}\right)^N\right\}$$

where $R = \prod_1^n A_i , R' = \prod_1^n (A_i + 2\delta)$, and $S_{n,r}$ is given by (2.9).

We now proceed to calculate $E(Y^2)$. For this purpose let $Q_1(y_1^1 , y_2^1 , \cdots , y_n^1)$ and $Q_2(y_1^2 , y_2^2 , \cdots , y_n^2)$ be two points which belong to R and $P_i(x_1^i , x_2^i , \cdots , x_n^i)$ be the centers of the N spheres $S_{n,r}$. Let us put

$$(4.3) \quad dQ_i = dy_1^i dy_2^i \cdots dy_n^i , (i = 1, 2), \quad dP_i = dx_1^i dx_2^i \cdots dx_n^i , (i = 1, 2, \cdots, N).$$

Consider the integral

$$(4.4) \qquad\qquad J = \int \frac{dQ_1 dQ_2 dP_1 dP_2 \cdots dP_N}{R'^N}$$

extended over the domain defined by

$$Q_1 \,\epsilon\, R, \quad Q_2 \,\epsilon\, R, \quad P_i \,\epsilon\, R', \quad \overline{Q_1 P_i} > r, \quad \overline{Q_2 P_i} > r, \qquad (i = 1, 2, \cdots, N).$$

If we keep $P_1 , P_2 , P_3 , \cdots , P_N$ fixed, each point Q_1 , Q_2 can vary independently over the set Y; consequently we have

$$(4.5) \qquad\qquad J = \int_{P_i \epsilon R'} \frac{Y^2 dP_1 dP_2 \cdots dP_N}{R'^N} = E(Y^2).$$

On the other hand, if we keep Q_1 and Q_2 fixed, the integral of each dP_i gives

$R' - 2 S_{n,r} + \mu(S_{n,r}, l)$ where $\mu(S_{n,r}, l)$ is given by (2.12) and $l = \overline{Q_1 Q_2}$. Hence we have

$$(4.6) \qquad J = \int_{Q_1 \epsilon R, Q_2 \epsilon R} \left(1 - \frac{2S_{n,r} - \mu(S_{n,r}, l)}{R'}\right)^N dQ_1\, dQ_2.$$

In order to calculate this integral we split it into two parts $J = J_1 + J_2$, according as $0 \leq l \leq 2r$ or $2r \leq l \leq D$, where $D = (\sum_1^n A_i^2)^{\frac{1}{2}}$. In the interval $0 \leq l \leq 2r$ we introduce the new variables of integration $l, \varphi_1, \varphi_2, \cdots, \varphi_{n-1}$ related to $y_1^1, y_2^1, \cdots y_n^1, y_1^2, y_2^2, \cdots, y_n^2$ by

$$(4.7) \qquad y_i^2 = y_i^1 + l\xi_i, \qquad\qquad (i = 1, 2, \cdots, n),$$

where ξ_i is given in (2.17). It is found that

$$\frac{\partial(y_1^1, y_2^1, \cdots, y_n^1, y_1^2, y_2^2, \cdots, y_n^2)}{\partial(y_1^1, y_2^1, \cdots, y_n^1, l, \varphi_1, \cdots \varphi_{n-1})} = l^{n-1} \sin^{n-2} \varphi_1 \sin^{n-3} \varphi_2 \cdots \sin \varphi_{n-2}.$$

Hence we have,

$$(4.8) \qquad dQ_1 dQ_2 = l^{n-1}\, dQ_1 d\sigma dl,$$

where $d\sigma$ denotes the element of area of the n-dimensional sphere of unit radius, given by (2.18). The same method used in section 3 gives

$$(4.9) \qquad J_1 = \int_0^{2r} \left(1 - \frac{2S_{n,r} - \mu(S_{n,r}, l)}{R'}\right)^N \Phi(R, l) l^{n-1}\, dl,$$

where $\Phi(R, l)$ is given by (2.23).

In the interval $2r \leq l \leq D$ $\mu(S_{n,r}, l) = 0$ and we have

$$(4.10) \qquad J_2 = \left(1 - \frac{2S_{n,r}}{R'}\right)^N \int_{2r \leq l \leq D} dQ_1 dQ_2 = \left(1 = \frac{2S_{n,r}}{R'}\right)^N$$
$$\cdot \left\{\int_{0 \leq l \leq D} dQ_1 dQ_2 - \int_{0 \leq l \leq 2r} dQ_1 dQ_2\right\}.$$

Now we have

$$(4.11) \qquad \int_{0 \leq l \leq D} dQ_1\, dQ_2 = R^2$$

and with the change of variables (4.7) we readily find that

$$(4.12) \qquad \int_{0 \leq l \leq 2r} dQ_1\, dQ_2 = \int_0^{2r} \Phi(R, l) l^{n-1}\, dl.$$

Collecting (4.9), (4.10), (4.11), (4.12) and taking into account (4.5) and (2.23) we have

$$E(Y^2) = \int_0^{2r} \left(1 - \frac{2S_{n,r} - \mu(S_{n,r}, l)}{R'}\right)^N \Phi(R, l) l^{n-1}\, dl$$

$$+ \left(1 - \frac{2S_{n,r}}{R'}\right)^N \left\{R^2 - \frac{2^n r^n}{n} R E_n - (-1)^n \frac{2^{3n} r^{2n}}{2n(2n-2)\cdots 4.2}\right.$$

(4.13)

$$- \sum_{k=1}^{n-1} (-1)^k \left(\sum_{i_1, i_2, \cdots, i_{n-k}} A_{i_1} A_{i_2} \cdots A_{i_{n-k}}\right)$$

$$\left.\cdot \frac{2^{n+2k} E_{n-k} r^{n+k}}{(n+k)(n+k-2)\cdots(n+k-2k)}\right\}$$

where $R = \prod_1^n A_i$, $R' = \prod_1^n (A_i + 2\delta)$; $S_{n,r}$ is given by (2.9), E_m by (2.20), $\mu(S_{n,r}, l)$ by (2.12) and $\Phi(R, l)$ by (2.23). In particular, for $n = 2$, we obtain the value given by Robbins [3, (30)], by use of (2.8), (2.15) and the equations $S_{2,n} = \pi r^2$, $E_i = 2$. For $n = 3$, the case of ordinary space it follows from (2.16), (2.24) and the equations $S_{3,r} = \frac{4}{3}\pi r^3$, $E_3 = 4\pi$, $E_2 = 2\pi$, that

$$E(Y^2) = \int_0^{2r} \left(1 - \frac{16\pi r^3 + 12\pi r^2 l - \pi l^3}{12R'} \cdots\right)^N \left(4\pi R - l^3 - 2\pi(A_1 A_2 + A_1 A_3\right.$$

$$+ A_2 A_3)l + \frac{8}{3}(A_1 + A_2 + A_3)l^2\right) l^2\, dl + \left(1 - \frac{8\pi r^3}{3R'}\right)^N \left\{R^2\right.$$

(4.14)

$$- \frac{32}{3}\pi R r^3 + 8\pi(A_1 A_3 + A_2 A_3 + A_2 A_3)r^4$$

$$\left. - \frac{256}{15}(A_1 + A_2 + A_3)r^5 + \frac{32}{3} r^6\right\}.$$

In this case the exact evaluation is easy if one expands the binomial under the sign of the integral and integrates term by term.

From (4.1) we see that $\sigma_X^2 = E(X^2) - E^2(X) = E(Y^2) - E^2(Y)$. Thus, from (4.2) and (4.13) we obtain immediately the second moment $E(X^2)$ and the variance σ_X^2 of X.

5. Remark. In the second problem we can substitute the n-intervals R and R' by concentric n-dimensional spheres. The problem may then be stated as follows:

Let $S_{n,a}$ denote a fixed n-dimensional sphere of radius a and $S_{n,a+\delta}$ the concentric n-dimensional sphere of radius $a + \delta$. $S_{n,a}$ and $S_{n,a+\delta}$ shall also denote the corresponding volumes. Let a fixed number N of n-dimensional spheres with radii r ($r \leq \min(a, \delta)$) be chosen independently with the probability density function for the center of each $S_{n,r}$ constant and equal to $1/S_{n,a+\delta}$ in $S_{n,a+\delta}$ and zero outside this n-sphere. Let X denote the intersection of the set-theoretical sum of the N n-spheres with $S_{n,a}$; we wish to evaluate the first two moments of the measure of X.

It suffices to observe that in this case we have

$$(5.1) \qquad \Phi(S_{n,a}, l) = \mu(S_{n,a}, l)E_n = 2a\ S_{n-1,a}E_n \int_0^\alpha \sin^n \theta\ d\theta$$

where $S_{n-1,a}$ is the volume of the $(n-1)$-dimensional sphere of radius a and $\alpha = \text{arc cos}\ (l/2a)$.

The same method used in section 4 gives

$$(5.2) \quad E(Y) = S_{n,a}\left(1 - \frac{S_{n,r}}{S_{n,a+\delta}}\right)^N, \qquad E(X) = S_{n,a}\left\{1 - \left(1 - \frac{S_{n,r}}{S_{n,a+\delta}}\right)^N\right\},$$

$$
\begin{aligned}
(5.3) \quad E(Y^2) = &\int_0^{2r}\left(1 - \frac{2S_{n,r} - \mu(S_{n,r}, l)}{S_{n,a+\delta}}\right)^N \Phi(S_{n,a}, l)l^{n-1}\ dl \\
&+ \left(1 - \frac{2S_{n,r}}{S_{n,a+\delta}}\right)^N\left\{S_{n,a}^2 - \int_0^{2r}\Phi(S_{n,a}, l)l^{n-1}\ dl\right\},
\end{aligned}
$$

where $\Phi(S_{n,a}, l)$ is given by (5.1).

In particular, for $n = 2$, by use of (5.1), (2.15) and the indefinite integrals

$$\int \text{arc cos}\ (l/2a)l\ dl = (\tfrac{1}{2}l^2 - a^2)\ \text{arc cos}\ (l/2a) - \tfrac{1}{4}\ l(4a^2 - l^2)^{\frac{1}{2}} + \text{constant},$$

$$\int l^2(4a^2 - l^2)^{\frac{1}{2}}\ dl = -\tfrac{1}{4}l(4a^2 - l^2)^{\frac{3}{2}} + \tfrac{1}{2}a^2l(4a^2 - l^2)^{\frac{1}{2}}$$

$$+ 2a^4\ \text{arc sin}\ (l/2a) + \text{constant}$$

we find that

$$
\begin{aligned}
E(Y^2) = 2\pi\int_0^{2r}&\left(1 - \frac{2\pi r^2 - 2r^2\ \text{arc cos}\ (l/2r) + \tfrac{1}{2}l(4r^2 - l^2)^{\frac{1}{2}}}{\pi(a + \delta)^2}\right)^N (2a^2\ \text{arc cos}(l/2a) \\
- \tfrac{1}{2}l(4a^2 - l^2)^{\frac{1}{2}}&\Big) l\ dl + \left(1 - \frac{2r^2}{(a + \delta)^2}\right)^N\left\{\pi^2 a^4 - 2\pi\left(2a^2(2r^2 - a^2)\ \text{arc cos}\left(\frac{r}{a}\right)\right.\right. \\
- 3a^2\ r(a^2 - r^2)^{\frac{1}{2}}& + \pi a^4 + 2r(a^2 - r^2)^{\frac{1}{2}} - a^4\ \text{arc sin}(r/a)\Big)\Big\}.
\end{aligned}
$$

For $n = 3$, we have by (5.1) and 2.16)

$$
\begin{aligned}
E(Y^2) = 4\pi\int_0^{2r}&\left(1 - \frac{16r^3 + 12r^2 l - l^3}{16(a + \delta)^3}\right)^N (\tfrac{4}{3}\pi a^3 - \pi a^2 l + \tfrac{1}{12}\pi l^3)l^2\ dl \\
&+ 4\pi\left(1 - \frac{2r^3}{(a + \delta)^3}\right)^N\left\{\tfrac{4}{3}\pi a^6 - \tfrac{3.2}{9}\pi a^3 r^3 + 4\pi a^2 r^4 - \tfrac{8}{9}\pi r^6\right\}.
\end{aligned}
$$

From (5.2) and (5.3) with the use of the relation $\sigma_X^2 = E(X^2) - E^2(X) = E(Y^2) - E^2(Y)$ we obtain immediately the second moment $E(X^2)$ and the variance σ_X^2 of X.

REFERENCES

[1] J. BRONOWSKI AND J. NEYMAN, "The variance of the measure of a two-dimensional random set," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 330–341.

[2] R. DELTHEIL, *Probabilitiés géométriques*, Gauthier-Villars, París, 1926.

[3] H. E. ROBBINS, "On the measure of a random set," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 70–74.

[4] H. E. ROBBINS, "On the measure of a random set, II." *Annals of Math. Stat.*, Vol. 16 (1945), pp. 342–347.

[5] L. A: SANTALÓ, "Sobre la medida cinemática en el plano", *Abhandlungen aus dem Mathematisches Seminar der Hamburgische Universität*, Vol. 11 (1936), pp. 222–236.

ON A TEST OF WHETHER ONE OF TWO RANDOM VARIABLES IS STOCHASTICALLY LARGER THAN THE OTHER

By H. B. Mann and D. R. Whitney

Ohio State University

1. Summary. Let x and y be two random variables with continuous cumulative distribution functions f and g. A statistic U depending on the relative ranks of the x's and y's is proposed for testing the hypothesis $f = g$. Wilcoxon proposed an equivalent test in the *Biometrics Bulletin*, December, 1945, but gave only a few points of the distribution of his statistic.

Under the hypothesis $f = g$ the probability of obtaining a given U in a sample of n x's and m y's is the solution of a certain recurrence relation involving n and m. Using this recurrence relation tables have been computed giving the probability of U for samples up to $n = m = 8$. At this point the distribution is almost normal.

From the recurrence relation explicit expressions for the mean, variance, and fourth moment are obtained. The $2r$th moment is shown to have a certain form which enabled us to prove that the limit distribution is normal if m, n go to infinity in any arbitrary manner.

The test is shown to be consistent with respect to the class of alternatives $f(x) > g(x)$ for every x.

2. Introduction. Let x and y be two random variables having continuous cumulative distribution functions f and g respectively. The variable x will be called stochastically smaller than y if $f(a) > g(a)$ for every a. We wish to test the hypothesis $f = g$ against the alternative that x is stochastically smaller than y. Such alternatives are of great importance in testing, for instance, the effect of treatments on some measurement. One may think of x as the values of certain measurements in the control group and of y as the values of the same measurement in a group which received treatment. In a particular instance the protective effect against infection by certain bacteria was investigated. Two groups of rats were used in the experiment. The first group receiving no treatment, the second group receiving the drug. Both groups were then infected with supposedly equally diluted cultures of the bacteria under investigation. Most of the rats in both groups died, but the time of survival was measured and it was desired to test whether the drug had the effect of prolonging the life of the rats. It was desired to make inferences from the effect on rats to the effect the drug would have on humans. Thus, the only relevant alternative to the hypothesis that survival times are not influenced by the drug is that the survival time of those rats which received treatment is stochastically larger than that of the control group.

3. The U test. Let the quantities $x_1, \cdots, x_n, y_1, \cdots, y_m$ be arranged in order. This arrangement is unique with probability 1 if $P(x_i = y_j) = 0$ and this follows from our assumption of continuity. Let U count the number of times a y precedes an x. If $P(U \leq \overline{U}) = \alpha$ under the null hypothesis, the test will be considered significant on the significance level α if $U \leq \overline{U}$ and the hypothesis of identical distributions of x and y will be rejected.

This test was first proposed by Wilcoxon [1]. His statistic T is the sum of the ranks of the y's in the ordered sequence of x's and y's. In general

$$U = mn + \frac{m(m + 1)}{2} - T$$

and this gives a simple way of computing U. Wilcoxon, however, treated only the case $m = n$ and in this case he tabulated only 3 points of the distribution of T. Since the test seems of great utility it seemed worthwhile to compute the variance, the moments and the limit distribution of U and to investigate the class of alternatives with respect to which the test is consistent.

Although this paper is written in terms of U and the probabilities of U are tabulated the results can be easily interpreted in terms of T if so desired.

4. The distribution of U. Consider now ordered sequences of n x's and m y's. Since it is only the relation between x and y that matters we replace each x by a 0 and each y by a 1. Let U count the number of times a 1 precedes a 0. Let $\bar{p}_{nm}(U)$ be the number of sequences of n 0's and m 1's in each of which a 1 precedes a 0 U times. By examining a sequence with the last term omitted we arrive at the recurrence relation:

$$\bar{p}_{nm}(U) = \bar{p}_{n-1m}(U - m) + \bar{p}_{nm-1}(U),$$

where $\bar{p}_{ij}(U) = 0$ if $U < 0$ and $\bar{p}_{i0}(U)$, $\bar{p}_{0i}(U)$ are zero or one according as $U \neq 0$ or $U = 0$.

Under the null hypothesis each of the $(m + n)!/m!n!$ sequences of n 0's and m 1's is equally likely. Consequently if $p_{nm}(U)$ represents the probability of a sequence in which a 1 precedes a 0 U times then

$$(1) \qquad p_{nm}(U) = \frac{n}{n + m} p_{n-1m}(U - M) + \frac{m}{n + m} p_{nm-1}(U).$$

Using the recurrence relation (1) the probabilities $p_{nm}(U)$ have been tabulated for $m \leq n \leq 8$ (see Table I). For $m = n = 8$ the distribution of $U - \frac{1}{2}(nm + 1)$ differs only a negligible amount from the normal distribution. We shall, in the following, derive the mean, the variance, and the fourth moment of U, and prove that the limit distribution of U is normal if n and m both approach infinity in any arbitrary manner.

It is obvious that $p_{nm}(U) = p_{mn}(U)$.

Since the probability of the ith 1 preceding the jth 0 is $\frac{1}{2}$, we have

$$(2) \qquad\qquad\qquad E_{nm}(U) = nm/2.$$

TABLE I

Probability of Obtaining a U not Larger than that Tabulated in Comparing Samples of n and m

n = 3

U \ m	1	2	2
0	.250	.100	.050
1	.500	.200	.100
2	.750	.400	.200
3		.600	.350
4			.500
5			.650

n = 4

U \ m	1	2	3	4
0	.200	.067	.028	.014
1	.400	.133	.057	.029
2	.600	.267	.114	.057
3		.400	.200	.100
4		.600	.314	.171
5			.429	.243
6			.571	.343
7				.443
8				.557

n = 5

U \ m	1	2	3	4	5
0	.167	.047	.018	.008	.004
1	.333	.095	.036	.016	.008
2	.500	.190	.071	.032	.016
3	.667	.286	.125	.056	.028
4		.429	.196	.095	.048
5		.571	.286	.143	.075
6			.393	.206	.111
7			.500	.278	.155
8			.607	.365	.210
9				.452	.274
10				.548	.345
11					.421
12					.500
13					.579

n = 6

U \ m	1	2	3	4	5	6
0	.143	.036	.012	.005	.002	.001
1	.286	.071	.024	.010	.004	.002
2	.428	.143	.048	.019	.009	.004
3	.571	.214	.083	.033	.015	.008
4		.321	.131	.057	.026	.013
5		.429	.190	.086	.041	.021
6		.571	.274	.129	.063	.032
7			.357	.176	.089	.047
8			.452	.238	.123	.066
9			.548	.305	.165	.090
10				.381	.214	.120
11				.457	.268	.155
12				.545	.331	.197
13					.396	.242
14					.465	.294
15					.535	.350
16						.409
17						.469
18						.531

TABLE I (*Continued*)

$n = 7$

U \ m	1	2	3	4	5	6	7
0	.125	.028	.008	.003	.001	.001	.000
1	.250	.056	.017	.006	.003	.001	.001
2	.375	.111	.033	.012	.005	.002	.001
3	.500	.167	.058	.021	.009	.004	.002
4	.625	.250	.092	.036	.015	.007	.003
5		.333	.133	.055	.024	.011	.006
6		.444	.192	.082	.037	.017	.009
7		.556	.258	.115	.053	.026	.013
8			.333	.158	.074	.037	.019
9			.417	.206	.101	.051	.027
10			.500	.261	.134	069	.036
11			.583	.324	.172	.090	.049
12				.394	.216	.117	.064
13				464	.265	.147	082
14				.538	.319	.183	.104
15					.378	.223	.130
16					.438	.267	.159
17					.500	.314	.191
18					.562	.365	.228
19						.418	.267
20						.473	.310
21						.527	.355
22							.402
23							.451
24							.500
25							.549

TABLE I (*Continued*)
$n = 8$

U	1	2	3	4	5	6	7	8	t	normal
0	.111	.022	.006	.002	.001	.000	.000	.000	3.308	.001
1	.222	.044	.012	.004	.002	.001	.000	.000	3.203	.001
2	.333	.089	.024	.008	.003	.001	.001	.000	3.098	.001
3	.444	.133	.042	.014	.005	.002	.001	.001	2.993	.001
4	.556	.200	.067	.024	.009	.004	.002	.001	2.888	.002
5		.267	.097	.036	.015	.006	.003	.001	2.783	.003
6		.356	.139	.055	.023	.010	.005	.002	2.678	.004
7		.444	.188	.077	.033	.015	.007	.003	2.573	.005
8		.556	.248	.107	.047	.021	.010	.005	2.468	.007
9			.315	.141	.064	.030	.014	.007	2.363	.009
10			.387	.184	.085	.041	.020	.010	2.258	.012
11			.461	.230	.111	.054	.027	.014	2.153	.016
12			.539	.285	.142	.071	.036	.019	2.048	.020
13				.341	.177	.091	.047	.025	1.943	.026
14				.404	.217	.114	.060	.032	1.838	.033
15				.467	.262	.141	.076	.041	1.733	.041
16				.533	.311	.172	.095	.052	1.628	.052
17					.362	.207	.116	.065	1.523	.064
18					.416	.245	.140	.080	1.418	.078
19					.472	.286	.168	.097	1.313	.094
20					.528	.331	.198	.117	1.208	.113
21						.377	.232	.139	1.102	.135
22						.426	.268	.164	.998	.159
23						.475	.306	.191	.893	.185
24						.525	.347	.221	.788	.215
25							.389	.253	.683	.247
26							.433	.287	.578	.282
27							.478	.323	.473	.318
28							.522	.360	.368	.356
29								.399	.263	.396
30								.439	.158	.437
31								.480	.052	.481
32								.520		

We now seek an expression for $E_{nm}(u^2)$ where $u = U - nm/2$. After multiplying (1) by $(U - nm/2)^2$, using

$$E_{nm}(u^2) = \sum_U (U - nm/2)^2 p_{nm}(U)$$

and expanding:

(3) $\qquad E_{nm}(u^2) = \dfrac{n}{n + m} E_{n-1m}(u^2) + \dfrac{m}{n + m} E_{nm-1}(u^2) + nm/4,$

where $E_{nm}(u)$ denotes the expectation of $(U - nm/2)$ in sequences with n 0's and m 1's. The initial conditions of (3) are seen by direct calculation to be

(4) $\qquad E_{n0}(u^2) = E_{0m}(u^2) = 0.$

By substitution $E_{nm}(u^2) = nm(n + m + 1)/12$ is a solution of the recurrence relation (3) and its initial conditions (4). Hence, it follows by mathematical induction that

(5) $\qquad E_{nm}(u^2) = nm(n + m + 1)/12.$

The fourth moment is similarly a solution of the recurrence relation

(6) $\quad E_{nm}(u^4) = \dfrac{n}{n + m} E_{n-1m}(u^4) + \dfrac{m}{n + m} E_{nm-1}(u^4)$

$$+ \dfrac{nm}{16}(2n^2m + 2nm^2 - n^2 - m^2 - nm)$$

which is obtained from (1) by multiplication by $(U - nm/2)^4$ and expansion. The initial conditions of (6) are found by direct calculation to be

(7) $\qquad E_{n0}(u^4) = E_{0m}(u^4) = 0.$

It may be verified that

(8) $E_{nm}(u^4) = \dfrac{nm(n + m + 1)}{240}(5n^2m + 5nm^2 - 2n^2 - 2m^2 + 3nm - 2n - 2m)$

satisfies the recurrence relation (6) and its initial conditions (7) and hence (8) follows by mathematical induction.

To investigate the limit distribution of u as n, m become infinite we investigate the rth moment. Following the same procedure as in the case of the second and fourth moments and using the symmetry of the distribution to find the odd moments zero we get the following recurrence relation.

(9) $\quad E_{nm}(u^{2r}) = \dfrac{1}{n + m}\sum_{\alpha=0}^{r}\binom{2r}{2\alpha}\dfrac{1}{4^{\alpha}}\{nm^{2\alpha}E_{n-1m}(u^{2r-2\alpha}) + mn^{2\alpha}E_{nm-1}(u^{2r-2\alpha})\}$

For $r = 1, 2$ it is known that $E_{nm}(u^{2r})$ is a polynomial in n and m of degree $3r$ and that it is divisible by $nm(n + m + 1)$. Assuming that $E_{nm}(u^{2\alpha})$, $\alpha < r$ is a polynomial in n and m of degree 3α divisible by $nm(n + m + 1)$ we will

show that it is possible to find a polynomial of degree $3r$ in n and m divisible by $nm(n + m + 1)$ which satisfies the recurrence relation (9) for $E_{nm}(u^{2r})$ and also its initial conditions, namely, $E_{n0}(u^{2r}) = E_{0m}(u^{2r}) = 0$.

The last condition is trivially satisfied if $E_{nm}(u^{2r})$ is divisible by $nm(n + m + 1)$. Our method here is to actually substitute a polynomial with undetermined coefficients into (9) and show that the coefficients can be obtained uniquely. Rearranging (9) we obtain

(10)
$$E_{nn}(u^{2r}) - \frac{n}{n + m} E_{n-1m}(u^{2r}) - \frac{m}{n + m} E_{nm-1}(u^{2r})$$
$$= \frac{1}{n + m} \sum_{\alpha=1}^{r} \binom{2r}{2\alpha} \frac{1}{4^{\alpha}} \{nm^{2\alpha} E_{n-1m}(u^{2r-2\alpha}) + mn^{2\alpha} E_{nm-1}(u^{2r-2\alpha})\}$$

Since for $\lambda < r$ we can write $E_{nm}(u^{2\lambda}) = nm(n + m + 1)P_{nm}^{3\lambda-3}$ where $P_{nm}^{3\lambda-3}$ is a polynomial in n, m of degree $3\lambda - 3$ the above equation reduces to

(11) $$E_{nm}(u^{2r}) - \frac{n}{n + m} E_{n-1m}(u^{2r}) - \frac{m}{n + m} E_{nm-1}(u^{2r}) = nmQ_{nm}^{3r-3}$$

where Q_{nm}^{3r-3} is a polynomial in n, m of degree $3r - 3$.

Now let

$$E_{nm}(u^{2r}) = nm(n + m + 1) \sum_{\substack{i,j=0 \\ i+j \le 3r-3}}^{3r-3} a_{ij} n^i m^j$$

where $a_{ij} = a_{ji}$ are to be determined. Substitution in (11) yields:

$$\sum_{i,j} a_{ij}[(n + m + 1)n^i m^j - (n - 1)(n - 1)^i m^j - (m - 1)n^i(m - 1)^j] = Q_{nm}^{3r-3}$$

and rearrangement yields:

(12) $$\sum_{\substack{i,j=0 \\ i+j \le 3r-3}}^{3r-3} a_{ij} \left[n^i m^j + \sum_{\alpha=0}^{i} \binom{i + 1}{\alpha} (-1)^{i-\alpha}(n^j m^\alpha + n^\alpha m^j) \right] = Q_{nm}^{3r-3}.$$

Consider first the terms of degree $3r - 3$. In this case $i + j = 3r - 3$ and $\alpha = i$ will give

$$\sum_{i=0}^{3r-3} a_{i3r-3-i}[n^i m^{3r-3-i} + (i + 1)(n^{3r-3-i} m^i + n^i m^{3r-3-i})]$$

or

(13) $$3r \sum_{i=0}^{3r-3} a_{i3r-3-i} n^i m^{3r-3-i}.$$

Equating the coefficients of these terms of degree $3r - 3$ to the corresponding ones in Q_{nm}^{3r-3} it is possible to calculate the value of $a_{i3r-3-i}$, $(i = 0, \cdots, 3r - 3)$.

We assume now that the a_{ij} are known for $i + j \ge 3r - 3 - (k - 1)$ and

we will find the value of a_{ij} where $i + j = 3r - 3 - k$.   Consider then the terms in (12) of degree $3r - 3 - k$.   These terms will occur when

$$i + j = 3r - 3, \alpha = i - k; i + j = 3r - 4, \alpha = i - k + 1; \cdots ;$$

$$i + j = 3r - 3 - k, \alpha = i.$$

All, but the last, contain coefficients which have already been evaluated.   The last one reduces to

$$(3r - 3) \sum_{i=0}^{3r-3-k} a_{i;3r-3-k-i}\, n^i\, m^{3r-3-k-i}.$$

Thus by equating coefficients $a_{i;3r-3-k-i}$ for $i = 0, 1, \cdots 3r - 3 - k$ can be evaluated in terms of the coefficients a_{ij} already known and those in Q_{nm}^{3r-3}. This concludes the proof that $E_{nm}(u^{2r})$ is a polynomial in n, m of degree $3r$ and is divisible by $nm(n + m + 1)$.

We now investigate the coefficients of the terms of degree $3r$.   For $\lambda = 1, 2$

$$E_{nm}(u^{2\lambda}) = \frac{(2\lambda - 1) \cdots 5\cdot3\cdot1}{12^\lambda}\, (nm)^\lambda(n + m + 1)^\lambda + \text{terms of degree} < 3\lambda).$$

We assume this to hold for $\lambda < r$ and we will show that it holds for $\lambda = r$.   Substitution reduces the right side of (10) to

$$\frac{1}{n + m} \binom{2r}{2} \frac{1}{4} \left\{ nm^2 \left[\frac{(2r - 3) \cdots 5\cdot3\cdot1}{12^{r-1}}\, (n - 1)^{r-1} m^{r-1}(n + m)^{r-1} \right] \right.$$

$$+ \left. mn^2 \left[\frac{(2r - 3) \cdots 5\cdot3\cdot1}{12^{r-1}}\, n^{r-1}(m - 1)^{r-1}(n + m)^{r-1} \right] + \text{(terms of degree} < 3r) \right\}$$

or

$$\frac{r(2r - 1)}{4} \left\{ (n + m)^{r-2} \left[\frac{(2r - 3) \cdots 5\cdot3\cdot1}{12^{r-1}} \right] \right.$$

$$\cdot [n(n - 1)^{r-1} m^{r+1} + m(m - 1)^{r-1} n^{r+1}] + \left. \text{(terms of degree} < 3r - 1) \right\}$$

which reduces to

$$\frac{3r(2r - 1) \cdots 5\cdot3\cdot1}{12^r}\, (nm)^r(n + m)^{r-1} + \text{(terms of degree} < 3r - 1).$$

Comparison of coefficients with (13) multiplied by nm gives

$$nm \sum_{i=0}^{3r-3} a_{i;3r-3-i}\, n^i\, m^{3r-3-i} = \frac{(2r - 1) \cdots 5\cdot3\cdot1}{12^r}\, (nm)^r(n + m)^{r-1}$$

or

(14) $$E_{nm}(u^{2r}) = \frac{(2r - 1) \cdots 5\cdot3\cdot1}{12^r}\, (nm)^r(n + m + 1)^r$$

$$+ \text{(terms of degree} < 3r).$$

We now wish to show that $E_{nm}(u^{2r})$ is at most of degree $2r$ in n or m. For $r = 1, 2$ this has already been established. Assuming that it is true for lower moments the right side of (10), which reduces to nmQ_{nm}^{3r-3} is at most of degree $2r - 1$ in n. We again compare coefficients in (12). First, for terms of degree $3r - 3$ we have already seen that n has degree at most $2r - 2$. For terms of degree $3r - 4$ we use $i + j = 3r - 3$, $\alpha = i - 1$ and $i + j = 3r - 4$, $\alpha = i$. The first case gives rise to no terms in n of degree greater than $2r - 2$ so when we solve for the coefficients $a_{i3r-4-i}$ the coefficients of terms in n of degree greater than $2r - 2$ must be zero. The process repeats and we find no terms in n or m of degree greater than $2r - 2$ in the left side of (12). This gives $E_{nm}(u^{2r})$ at most the degree $2r$ in n or m.

Now consider the ratio

$$I \equiv \frac{E_{nm}(u^{2r})}{[E_{nm}(u^2)]^r}$$

$$= \frac{\frac{(2r - 1) \cdots 5 \cdot 3 \cdot 1}{12^r} (nm)^r (n + m + 1)^r}{[nm(n + m + 1)/12]^r}$$

$$+ \frac{\text{(terms of degree} < 3r; \text{ in } n \text{ or } m, \leq 2r)}{[nm(n + m + 1)/12]^r}$$

$$= (2r - 1) \cdots 5 \cdot 3 \cdot 1 + \frac{\text{(terms of degree} < 3r; \text{ in } n \text{ or } m, \leq 2r)}{(nm)^r (n + m + 1)^r} .$$

Hence

(15) $$\operatorname*{Lim}_{n,m\to\infty} I = (2r - 1) \cdots 5 \cdot 3 \cdot 1$$

and by a well known theorem it follows from (15) that the limit distribution is normal.

5. Consistency of the U test. If f and g are the cumulative distribution functions of the x's and y's then our null hypothesis is $f = g$. The alternatives admitted are $f(a) > g(a)$ for every a. Let E_A denote the expectation under the alternative.

Defining

$$x_{ij} = \begin{matrix} 0 \text{ if } x_i < y_j \\ 1 \text{ if } x_i > y_j \end{matrix}$$

we have

$$E_A(x_{ij}) = P(x_i > y_j) = \int_{-\infty}^{\infty} g \, df < \tfrac{1}{2}$$

$$E_A(x_{ij}x_{ik}) = P(x_i > y_j ; x_i > y_k) = \int_{-\infty}^{\infty} g^2 \, df < \tfrac{1}{3}$$

$$E_A(x_{ik}x_{jk}) = P(x_i > y_k, x_j > y_k) = \int_{-\infty}^{\infty} (1 - f)^2 \, dg < \tfrac{1}{3}.$$

We can now write

$$E_A(x_{ij}) = \tfrac{1}{2} - \lambda, \qquad E_A(x_{ij}x_{ik}) = \tfrac{1}{3} - \epsilon_1, \qquad E_A(x_{ik}x_{jk}) = \tfrac{1}{3} - \epsilon_2$$

where λ, ϵ_1, ϵ_2 are positive numbers.
We have then

$$\sigma_A^2(x_{ij}) = \tfrac{1}{4} - \lambda^2 \qquad\qquad \sigma_A(x_{ij}x_{ik}) = \tfrac{1}{12} - \epsilon_1 + \lambda - \lambda^2$$

$$\sigma_A(x_{ij}x_{kl}) = 0 \text{ for } i \neq k, j \neq l \quad \sigma_A(x_{ik}x_{jk}) = \tfrac{1}{12} - \epsilon_2 + \lambda - \lambda^2$$

Now

(16)
$$E_A(U) = \sum_{i,j} E_A(x_{ij}) = nm/2 - \lambda nm$$

and

(17) $\quad \sigma_A^2(U) = \Sigma\, \sigma_A^2(x_{ij}) + \Sigma\, \sigma_A(x_{ij}x_{ik}) + \Sigma\, \sigma_A(x_{ik}x_{jk}) + \Sigma\, \sigma_A(x_{ij}x_{kl})$

or

$$\sigma_A^2(U) = nm(n + m + 1)/12$$
$$+ nm[-\lambda^2(n + m - 1) + (\lambda - \epsilon_1)(m - 1) + (\lambda - \epsilon_2)(n - 1)].$$

Let the critical region under the null hypothesis consist of those U's satisfying $nm/2 - U \geq t_n\sigma$ where $\lim_{n\to\infty} t_n = t$. Then

$$P(nm/2 - U \geq t_n\sigma\,|\,A) = P(E_A(U) - U \geq k\cdot\sigma_A) \quad\text{where}\quad k = \frac{t_n\sigma - \lambda nm}{\sigma_A}$$

and by Tchebycheff's inequality, since for large values of n,m $k < 0$

$$P(nm/2 - U \geq t_n\sigma\,|\,A) \geq 1 - \frac{\sigma_A^2}{(t_n\sigma - \lambda nm)^2},$$

which by (5) and (17) gives

$$P(nm/2 - U \geq t_n\sigma\,|\,A) \geq 1$$
$$- \frac{\dfrac{nm(n + m + 1)}{12} + nm[-\lambda^2(n + m - 1) + (\lambda - \epsilon_1)(m - 1) + (\lambda - \epsilon_2)(n - 1)]}{(t_n\sqrt{nm(n + m + 1)/12} - \lambda nm)^2}$$

$$\geq 1$$
$$- \frac{1 + \dfrac{12}{n + m + 1}[-\lambda^2(n + m - 1) + (\lambda - \epsilon_1)(m - 1) + (\lambda - \epsilon_2)(n - 1)]}{\left(t_n - \lambda\sqrt{\dfrac{12nm}{n + m + 1}}\right)^2}.$$

We obtain then that

$$\operatorname{Lim}_{n,m\to\infty} P(nm/2 - U \geq t_n\sigma\,|\,A) = 1$$

which is the requirement for consistency.

6. Comparison with other tests. Another test which might seem appropriate for the comparison of a control group with a group receiving treatment is the test introduced by Wald and Wolfowitz [2]. The test by Wald and Wolfowitz is consistent with respect to every alternative g. However in the case considered we are only interested in the alternative hypothesis that measurements in the group receiving treatment are stochastically larger than in the control group. Intuitively, it seems that the test proposed here is more efficient for detecting the particular alternative considered than the test proposed by Wald and Wolfowitz. This intuitive feeling was borne out by the results of the test in the particular experiment described in the introduction. All in all, 62 experiments were conducted using various bacteria in different solutions and various amounts of the protective drug. The U Test gave 14 significant results on the 5% level and 4 on the 1% level. The test of Wald and Wolfowitz gave 7 significant results on the 5% level and 2 on the 1% level. A final decision between the two tests can, of course, only be arrived at on the basis of their power functions, which present formidable difficulties.

In comparing the two statistics it was noted that a slight dislocation of a value may cause a significant change in the number of runs easier than it can cause a significant change in the statistic proposed here. For instance, in the sequence $x_1x_2x_3x_4x_5x_6y_1y_2y_3y_4y_5y_6$ both statistics would give a probability less than .05. If however, the sequence is slightly altered to $x_1x_2x_3x_4x_5y_1x_6y_2y_3y_4y_5y_6$, P (number of runs ≤ 4) $> .05$ while $P(U \leq 1) = .002$.

After completion of the present paper it came to the authors attention that the U test had already been proposed by K. K. Mathen [3]. However Mathen's distribution of U is incorrect and its derivation erroneous, since it assumes independence of the random variables x_{ij} as defined in section 5 of the present paper, while obviously x_{ij} and x_{ik} are not independent.

REFERENCES

[1] FRANK WILCOXON, "Individual comparisons by ranking methods", *Biometrics Bull.*, Vol. 1 (1945), pp. 80–83.
[2] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population", *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.
[3] K. K. MATHEN, *Sankhyā*, 1946, p. 329.

ON THE CONVERGENCE OF SEQUENCES OF MOMENT GENERATING FUNCTIONS

By W. Kozakiewicz

University of Saskatchewan

1. Summary. The purpose of this paper is to give a few theorems concerning the reciprocal relation between the convergence of a sequence of distribution functions and the convergence of the corresponding sequence of their moment generating functions.

The paper consists of two parts. In the first part the univariate case is discussed. The content of this part is closely related to that of a recent paper by J. H. Curtiss [1, p. 430–433], but the results are of a somewhat more general nature, and the methods of proofs are different and do not make use of the theory of a complex variable. The second part deals with the multivariate case which, as far as the author knows, has not been treated before with proofs in as complete and rigorous a way.

In both the univariate and multivariate cases the proofs are based on the well known Helly selection principle [2, p. 26] for bounded sequences of monotonic functions.

2. The univariate case. Let X be a random variable and $F(x)$ its distribution function. That is, for any real x, $F(x) = P\{X \leq x\}$, where $P\{X \leq x\}$ denotes the probability of the event $X \leq x$. The function

$$\varphi(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tX} \, dF(x),$$

in which the integral is taken in the Stieltjes-Riemann sense and is assumed to converge in some neighborhood of the origin, is called the moment generating function of X (or of $F(x)$).

Henceforth we use the abbreviations d.f. and m.g.f. for the terms distribution function and moment generating function respectively. The variable t will be always real.

THEOREM 1. *Let $\{F_n(x)\}$ be a sequence of d.f.'s. Let $M(x)$ for any fixed non-negative x be the least upper bound of the sequence $\{F_n(-x) + 1 - F_n(x)\}$. If the sequence $\{F_n(x)\}$ converges on an everywhere dense set of points on the x-axis, and if there exists a positive number α such that for any fixed t in the interval $|t| < \alpha$*

$$(1) \qquad \lim_{x \to +\infty} e^{|t|x} M(x) = 0,$$

then:

(a) *there exists a d.f. $F(x)$ such that $\lim_{n \to \infty} F_n(x) = F(x)$ at each point of continuity of of $F(x)$;*

(b) *the m.g.f.'s of $F(x)$ and $F_n(x)$, say $\varphi(t)$ and $\varphi_n(t)$ exist for $|t| < \alpha$;*

(c) *$\lim_{n \to \infty} \varphi_n(t) = \varphi(t)$ for $|t| < \alpha$ and uniformly in each interval $|t| \leq \beta < \alpha$.*

To prove (a), it may be noticed that there exists a function $F(x)$, non-decreasing and continuous on the right, such that $\lim_{n \to \infty} F_n(x) = F(x)$ at each point of continuity of $F(x)$. But $F(x)$ must be a distribution function. Indeed, we have for $x > 0$

$$(2) \qquad\qquad F(-x) + 1 - F(x) \leq M(x-).$$

Now from (1), putting $t = 0$, we find that $M(x)$ and consequently $M(x-)$ approach zero as $x \to +\infty$. This proves that $F(-\infty) = 0$ and $F(+\infty) = 1$.

To prove (b), we notice first that the integral

$$\varphi_n(t) = \int_{-\infty}^{+\infty} e^{xt}\, dF_n(x) \qquad\qquad (n = 1, 2, \cdots),$$

is convergent for $|t| < \alpha$. This follows immediately from (1) by applying the method of integration by parts to the integrals

$$\int_0^N e^{xt}\, dF_n(x) \quad \text{and} \quad \int_{-N}^0 e^{xt}\, dF_n(x),$$

which for any t in the interval $|t| < \alpha$ will be seen to be bounded for all values of N. By the same argument, the relation $\lim_{x \to +\infty} M(x-)e^{|t||x|} = 0$, $|t| < \alpha$, which can be easily deduced from (1), together with (2) imply that the integral representing $\varphi(t)$ is convergent for $|t| < \alpha$.

Let now β be a positive number less than α and let γ be such that $\beta < \gamma < \alpha$. Let M_γ be the least upper bound of $M(x)e^{\gamma x}$ for $x > 0$. Using the method of integration by parts and applying (1) we have for $|t| \leq \beta$

$$(3) \qquad \begin{aligned} \int_N^{+\infty} e^{xt}\, dF_n(x) &= [1 - F_n(N)]\, e^{Nt} + t \int_N^{+\infty} e^{xt}[1 - F_n(x)]\, dx \\ &\leq M(N)e^{N\beta} + M_\gamma \beta\, \frac{e^{N(\beta-\gamma)}}{\gamma - \beta}. \end{aligned}$$

We could prove easily that the same inequality is true for the integrals

$$\int_{-\infty}^{-N} e^{xt}\, dF_n(x), \qquad \int_N^{+\infty} e^{xt}\, dF(x), \qquad \int_{-\infty}^{-N} e^{xt}\, dF(x).$$

Now let ϵ be any positive number. Because of (3), we have

$$(4) \qquad \int_{|x| > N_0} e^{xt}\, dF_n(x) < \epsilon, \qquad \int_{|x| > N_0} e^{xt}\, dF(x) < \epsilon,$$

for a sufficiently great number N_0, and uniformly with respect to n and t, when $|t| \leq \beta$. Clearly, N_0 can be so chosen that $F(x)$ is continuous for $x = \pm N_0$. Then

$$(5) \qquad \lim_{n \to \infty} \int_{-N_0}^{N_0} e^{xt}\, dF_n(x) = \int_{-N_0}^{N_0} e^{xt}\, dF(x),$$

uniformly for $|t| \leq \beta$.

The relations (4) and (5) prove that $\varphi_n(t) \to \varphi(t)$ as $n \to \infty$, uniformly for $|t| \leq \beta$. But β can be chosen as near to α as we please; thus (c) is proved.

THEOREM 2. *Let $\{F_n(x)\}$ be a sequence of d.f.'s and $\{\varphi_n(t)\}$ the corresponding sequence of m.g.f.'s. If $\varphi_n(t)$ exists for $|t| < \alpha$, and if there exists a finite valued function $\varphi(t)$ defined for $|t| < \alpha$, such that $\lim \varphi_n(t) = \varphi(t)$ for $|t| < \alpha$, then*

(a)
$$\lim_{x \to +\infty} M(x)e^{|t|x} = 0 \ for \ |t| < \alpha;$$

(b) *there exists a d.f. $F(x)$ such that $\lim_{n \to \infty} F_n(x) = F(x)$ at each point of continuity of $F(x)$*

(c) *the m.g.f. of $F(x)$ exists for $|t| < \alpha$ and is identically equal to $\varphi(t)$ in this interval.*

(d) $\lim_{n \to \infty} \varphi_n(t) = \varphi(t)$ *uniformly in each interval $|t| \leq \beta < \alpha$.*

To prove (a), let t be a number in the interval $|t| < \alpha$, and let β be chosen so that $|t| < \beta < \alpha$. Then, for $x \geq 0$, we have

$$F_n(-x) + 1 - F_n(x) = \int_{-\infty}^{-x} dF_n(u) + \int_{x}^{+\infty} dF_n(u)$$

$$\leq e^{-\beta x} \int_{-\infty}^{-x} e^{-\beta u} \, dF_n(u) + e^{-\beta x} \int_{x}^{+\infty} e^{\beta u} \, dF_n(u)$$

$$\leq e^{-\beta x} [\varphi_n(-\beta) + \varphi_n(\beta)].$$

Consequently

$$M(x)e^{|t|x} \leq e^{(|t|-\beta)x} \, \text{l.u.b.} \{\varphi_n(-\beta) + \varphi_n(\beta)\},$$

and since the sequences $\{\varphi_n(-\beta)\}$ and $\{\varphi_n(\beta)\}$ are convergent, and therefore bounded, it follows that $M(x)e^{|t|x}$ approaches zero as $x \to +\infty$.

To prove (b) we may notice that by the Helly selection principle we can choose a subsequence $\{F_{n_k}(x)\}$ which is convergent to some non-decreasing function $F(x)$, at each point of continuity of $F(x)$. Now the Theorem 1 together with (a) imply that $F(x)$ is a d.f. and that the limit of the subsequence $\{\varphi_{n_k}(t)\}$, namely $\varphi(t)$, must be identical, for $|t| < \alpha$, with the m.g.f. of $F(x)$. By the uniqueness property of a m.g.f. we know that $F(x)$ is uniquely determined by $\varphi(t)$, and therefore it follows that every convergent subsequence of $\{F_n(x)\}$ approaches the same limit $F(x)$ at each point of continuity of $F(x)$. This is, however, equivalent to the statement that the sequence $\{F_n(x)\}$ itself converges to $F(x)$ at each point of continuity of $F(x)$. Thus (b) is proved. We see at once that (c) and (d) follow immediately from the Theorem 1.

Theorem 2 is of course similar to the Theorem 3 in the paper of Curtiss [1, p. 432]. The proof of (a), however, is not contained in his paper. From the Theorems 1 and 2 there follows immediately

THEOREM 3. *Let $\{F_n(x)\}$ be a sequence of d.f.'s, and let $\{\varphi_n(t)\}$ be the correspond-*

ing sequence of m.g.f.'s, which are all assumed to exist for $|t| < \alpha$. The necessary and sufficient conditions for the convergence of $\{\varphi_n(t)\}$ in the interval $|t| < \alpha$, are:

(a) $\lim\limits_{n \to +\infty} M(x)e^{|t|x} = 0,$ $|t| < \alpha$

(b) *the sequence $\{F_n(x)\}$ converges to a d.f. $F(x)$ at each point of continuity of $F(x)$. Further, the m.g.f. of $F(x)$ exists for $|t| < \alpha$ and is equal in this interval to the limit of the sequence $\{\varphi_n(t)\}$.*

In his paper Curtiss gives an example of a sequence $\{F_n(x)\}$ of d.f.'s which converges to a d.f. $F(x)$, while the corresponding sequence $\{\varphi_n(t)\}$ of m.g.f.'s does not converge to the m.g.f. $\varphi(t)$ of the d.f. $F(x)$, though both $\varphi_n(t)$, $(n = 1, 2, \cdots)$, and $\varphi(t)$ exist for all t. It may be easily proved by the direct method that in the case considered the condition (a) of the Theorem 3 is not satisfied.

It is perhaps worth while to notice that the condition (a) of the Theorem 3 may be expressed also as follows:

$$\varlimsup_{x \to +\infty} x^{-1} \log M(x) \leq -\alpha.$$

3. The multivariate case. For the sake of simplicity we shall consider here the bivariate case only. The results obtained in this chapter, can be, however, easily extended to the case when d.f.'s and m.g.f.'s are defined in the Euclidean space of any finite number of dimensions.

Let (X_1, X_2) be a random vector variable in the two-dimensional Euclidean space, and let $F(x_1, x_2)$ be its d.f. That is, for any real numbers x_1 and x_2,

$$F(x_1, x_2) = P\{X_1 \leq x_1, X_2 \leq x_2\}.$$

Let

$$F_1(x_1) = P\{X_1 \leq x_1\} = F(x_1, +\infty),$$

$$F_2(x_2) = P\{X_2 \leq x_2\} = F(+\infty, x_2);$$

then $F_1(x_1)$ and $F_2(x_2)$ are called the marginal d.f.'s of X_1 and X_2 respectively. The m.g.f.'s of the d.f.'s $F(x_1, x_2)$, $F_1(x_1)$ and $F_2(x_2)$ are defined by the equations:

$$\varphi(t_1, t_2) = E(e^{X_1 t_1 + X_2 t_2}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{x_1 t_1 + x_2 t_2} \, dF(x_1, x_2)$$

$$\varphi_i(t_i)) = E(e^{X_i t_i}) = \int_{-\infty}^{+\infty} e^{x_i t_i} \, dF_i(x_i), \qquad (i = 1, 2),$$

in which the integrals are assumed to converge in some neighborhood of the origin. It is easy to see that $\varphi_1(t_1) = \varphi(t_1, 0)$ and $\varphi_2(t_2) = \varphi(0, t_2)$.

THEOREM 4. *Let $\varphi(t_1, t_2)$ and $\varphi^*(t_1, t_2)$ be the m.g.f.'s of d.f.'s $F(x_1, x_2)$ and $F^*(x_1, x_2)$ respectively. If $\varphi(t_1, t_2)$ and $\varphi^*(t_1, t_2)$ exist and are equal in some neighborhood of the origin $|t_i| < \alpha_i$, $(i = 1, 2)$, then $F(x_1, x_2) = F^*(x_1, x_2)$ identically.*

To prove this theorem, let us introduce two random vector variables (X_1, X_2)

and (X_1^*, X_2^*) of which the d.f.'s are respectively F and F^*. Consider now two random variables

$$Z = X_1 t_1 + X_2 t_2, \qquad Z^* = X_1^* t_1 + X_2^* t_2,$$

where t_1 and t_2 denote two real numbers not both zero. If $\varphi(t)$ and $\varphi^*(t)$ are respectively the m.g.f.'s of Z and Z^*, we have

$$\varphi(t) = \varphi(t t_1, t t_2), \qquad \varphi^*(t) = \varphi^*(t t_1, t t_2).$$

Consequently $\varphi(t) = \varphi^*(t)$ provided that $|t t_i| < \alpha_i, (i = 1, 2)$. It follows from the uniqueness property of the m.g.f. in the univariate case that the d.f.'s of Z and Z^* must be identical. Now, according to a theorem due to Cramér [3, p. 105], if the d.f.'s of Z and Z^* coincide for all pairs of values (t_1, t_2) such that $|t_1| + |t_2| \neq 0$, the d.f.'s F and F^* must be identical. It may be worth while to reproduce here Cramér's proof. Let $\psi(t_1, t_2) = E(e^{i(X_1 t_1 + X_2 t_2)})$ and $\psi^*(t_1, t_2) = E(e^{i(X_1^* t_1 + X_2^* t_2)})$ be the characteristic functions of F and F^* respectively. Then $\psi(t t_1, t t_2)$ and $\psi^*(t t_1, t t_2)$ are the characteristic functions of Z and Z^* respectively. Since Z and Z^* have the same d.f.'s, it follows that $\psi(t t_1, t t_2) = \psi^*(t t_1, t t_2)$ for all values of t. Putting $t = 1$, we find that $\psi(t_1, t_2) = \psi^*(t_1, t_2)$ if $|t_1| + |t_2| \neq 0$. For $t_1 = t_2 = 0, \psi(0, 0) = \psi^*(0, 0) = 1$. Therefore $\psi(t_1, t_2) = \psi^*(t_1, t_2)$ identically, and since the characteristic function uniquely determines the d.f., it follows that the d.f. F and F^* are identical.

THEOREM 5. *Let $\{F_n(x_1, x_2)\}$ be a sequence of d.f.'s. Let $F_{1n}(x_1)$ and $F_{2n}(x_2)$ be respectively the marginal d.f.'s determined by $F_n(x_1, x_2)$. Let*

$$M_i(x_i) = \text{l.u.b.}_n \{F_{in}(-x_i) + 1 - F_{in}(x_i)\}$$

where $x_i \geq 0, (i = 1, 2)$. If there exist positive numbers α_1 and α_2 such that for $|t_i| < \alpha_i$

$$(6) \qquad \lim_{x_i \to +\infty} M_i(x_i) e^{|t_i| x_i} = 0, \qquad (i = 1, 2),$$

and if $\{F_n(x_1, x_2)\}$ converges on an everywhere dense set on the (x_1, x_2) plane, then:

(a) there exists a d.f. $F(x_1, x_2)$ such that $\lim_{n \to \infty} F_n(x_1, x_2) = F(x_1, x_2)$ at each point of continuity of $F(x_1, x_2)$,

(b) there exist two positive numbers δ_1 and $\delta_2, \delta_i < \alpha_i$, such that the m.g.f.'s of $F(x_1, x_2)$ and $F_n(x_1, x_2)$, say $\varphi(t_1, t_2)$ and $\varphi_n(t_1, t_2)$, exist for $|t_i| < \delta_i, (i = 1, 2)$,

(c) $\lim_{n \to \infty} \varphi_n(t_1, t_2) = \varphi(t_1, t_2)$ for $|t_i| < \delta_i$, and uniformly in each two-dimensional interval $|t_i| \leq \beta_i < \delta_i, (i = 1, 2)$.

To prove (a), we notice that there obviously exists a function $F(x_1, x_2)$, continous on the right with respect to each variable, satisfying the relation

$$\Delta^2 F(x_1, x_2) = F(x_1'', x_2'') + F(x_1', x_2') - F(x_1', x_2'') - F(x_1'', x_2') \geq 0$$

for $x_1' \leq x_1'', x_2' \leq x_2''$, and such that

(7) $$\lim_{n \to \infty} F_n(x_1, x_2) = F(x_1, x_2)$$

at each point of continuity of $F(x_1, x_2)$. We shall prove that $F(x_1, x_2)$ is a d.f. In fact, it is easy to see that we have for $x_i > 0$, $(i = 1, 2)$,

(8) $$F(-x_1, -x_2) \leq F(-x_1, x_2) \leq M_1(x_1-), \quad F(x_1, -x_2) \leq M_2(x_2-),$$
$$1 - F(x_1, x_2) \leq M_1(x_1) + M_2(x_2).$$

Now, according to (6), $\lim_{x_i \to +\infty} M_i(x_i-) = \lim_{x_i \to +\infty} M_i(x_i) = 0$, $(i = 1, 2)$, therefore it follows from (8) that $F(-\infty, -\infty) = F(-\infty, x_2) = F(x_1, -\infty) = 0$ and $F(+\infty, +\infty) = 1$, which proves that $F(x_1, x_2)$ is a d.f.

To prove (b), let $\varphi_{in}(t_i)$ be the m.g.f. of the d.f. $F_{in}(x_i)$, $(i = 1, 2)$. Let $F_1(x_1)$ and $F_2(x_2)$ be the marginal d.f.'s determined by $F(x_1, x_2)$ and let $\varphi_i(t_i)$ be the m.g.f. of $F_i(x_i)$, $(i = 1, 2)$.

Now let $N' > N > 0$ and

$$R_n(N, N', t_1, t_2) = \int_{-N'}^{N'} \int_{-N'}^{N'} e^{x_1 t_1 + x_2 t_2} dF_n(x_1, x_2) - \int_{-N}^{N} \int_{-N}^{N} e^{x_1 t_1 + x_2 t_2} dF_n(x_1, x_2)$$

$$= \int_{N}^{N'} \int_{-N'}^{N} + \int_{-N}^{N'} \int_{N}^{N'} + \int_{-N'}^{-N} \int_{-N}^{N'}$$

$$+ \int_{-N'}^{N} \int_{-N'}^{-N} e^{x_1 t_1 + x_2 t_2} dF_n(x_1, x_2) = I_1 + I_2 + I_3 + I_4.$$

Applying the Schwartz inequality to I_1, we find

(9) $$I_1 \leq \left(\int_{N}^{N'} \int_{-N'}^{N} e^{2x_1 t_1} dF_n \right)^{\frac{1}{2}} \left(\int_{N}^{N'} \int_{-N'}^{N} e^{2x_2 t_2} dF_n \right)^{\frac{1}{2}}.$$

But

(10) $$\int_{N}^{N'} \int_{-N'}^{N} e^{2x_1 t_1} dF_n(x_1, x_2) \leq \int_{N}^{N'} e^{2x_1 t_1} dF_{1n}(x_1),$$

and similarly

(11) $$\int_{N}^{N'} \int_{-N'}^{N} e^{2x_2 t_2} dF_n(x_1, x_2) \leq \int_{-N'}^{N} e^{2x_2 t_2} dF_{2n}(x_2).$$

Let ϵ be any positive number and γ_i a positive number less than α_i, $(i = 1, 2)$. It follows from the proof of the Theorem 1, taking into account (6), that the integrals representing $\varphi_{in}(t_i)$ and $\varphi_i(t_i)$, $(i = 1, 2)$, exist and are uniformly convergent with respect to n and t_i, when $|t_i| \leq \gamma_i$, $(i = 1, 2)$. Consequently we have

(12) $$\int_{|x_i| > N} e^{x_i t} dF_{in}(x_i) < \epsilon, \qquad \int_{|x_i| > N} e^{x_i t_i} dF_i(x_i) < \epsilon, \qquad (i = 1, 2),$$

uniformly with respect to n and t_i when $|t| \leq \gamma_i$, $(i = 1, 2)$, provided that N is sufficiently large, say $N \geq N_0$. Let us take $\beta_i = \gamma_i/2$, $(i = 1, 2)$. The integrals

representing $\varphi_{in}(t_i)$ and $\varphi_i(t_i)$, $(i = 1, 2)$, are obviously uniformly bounded for all n and when $|t_i| \leq \gamma_i$, $(i = 1, 2)$, they are all less than some constant C. Consequently taking into account (9), (10), (11), and (12), we find

$$I_1 < \sqrt{C\epsilon},$$

uniformly with respect to n and t_i when $|t_i| \leq \beta_i$, $(i = 1, 2)$, provided that $N' > N \geq N_0$. Since the same inequality is true for I_2, I_3 and I_4, we have

(13) $$R_n(N, N', t_1, t_2) < 4\sqrt{C\epsilon},$$

uniformly with respect to n and t_i, when $|t_i| \leq \beta_i$, $(i = 1, 2)$, provided $N' > N \geq N_0$. Hence the integral representing $\varphi_n(t_1, t_2)$ is uniformly convergent for $|t_i| \leq \beta_i$, and consequently convergent for $|t_i| < \alpha_i/2$, $(i = 1, 2)$, since β_i can be chosen as near to $\alpha_i/2$ as we please.

Similarly, using (12), we could find

(14) $$R(N, N', t_1, t_2) < 4\sqrt{C\epsilon}, \qquad |t_i| \leq \beta_i, N' > N \geq N_0$$

where

$$R(N, N', t_1, t_2) = \int_{-N'}^{N'} \int_{-N'}^{N'} e^{x_1 t_1 + x_2 t_2} dF(x_1, x_2) - \int_{-N}^{N} \int_{-N}^{N} e^{x_1 t_1 + x_2 t_2} dF(x_1, x_2).$$

This proves, in turn, that the integral representing $\varphi(t_1, t_2)$ is uniformly convergent for $|t_i| \leq \beta_i$ and convergent for $|t_i| < \alpha_i/2$, $(i = 1, 2)$. Thus (b) is proved with $\delta_i = \alpha_i/2$, $(i = 1, 2)$.

To prove (c), let $N' \to +\infty$ and $N = N_0$ in (13) and (14). We obtain

(15) $$R_n(N_0, +\infty, t_1, t_2) \leq 4\sqrt{C\epsilon}, \qquad R(N_0, +\infty, t_1, t_2) \leq 4\sqrt{C\epsilon}$$

uniformly with respect to n and t_i when $|t_i| \leq \beta_i$.

Clearly, N_0 can be chosen so that $F_1(x_1)$ and $F_2(x_2)$ are continuous for $x_1 = x_2 = \pm N_0$. Then

(16) $$\lim_{n \to \infty} \int_{-N_0}^{N_0} \int_{-N_0}^{N_0} e^{x_1 t_1 + x_2 t_2} dF_n(x_1, x_2) = \int_{-N_0}^{N_0} \int_{-N_0}^{N_0} e^{x_1 t_1 + x_2 t_2} dF(x_1, x_2),$$

uniformly for $|t_i| \leq \beta_i$, $(i = 1, 2)$.

The relations (15) and (16) prove that

$$\lim_{n \to \infty} \varphi_n(t_1, t_2) = \varphi(t_1, t_2),$$

uniformly for $|t_i| \leq \beta_i$, $(i = 1, 2)$. The ordinary convergence obviously holds for $|t_i| < \alpha_i/2$, $(i = 1, 2)$.

It follows from the above proof, which refers to the bivariate case, that we may take $\delta_i = \alpha_i/2$, $(i = 1, 2)$, in (b) and (c).

The existence of the corresponding numbers δ_1, $\delta_i < \alpha_i$, $(i = 1, 2, \cdots, k)$, in the k-variate case can be easily established by the repeated application of the Schwartz inequality.

THEOREM 6. *Let* $\varphi_n(t_1, t_2)$, $\varphi_{in}(t_i)$, $F_n(x_1, x_2)$, $F_{in}(x_i)$ *and* $M_i(x_i)$, $(i = 1, 2)$, *have the same meaning as in the Theorem 5. If* $\varphi_n(t_1, t_2)$ *exist for* $|t_i| < \alpha_i$, $(i = 1, 2)$, *and if there exists a finite valued function* $\varphi(t_1, t_2)$ *defined for* $|t_i| < \alpha_i$, *such that* $\lim_{n \to \infty} \varphi_n(t_1, t_2) = \varphi(t_1, t_2)$, $|t_i| < \alpha_i$,

then

(a)
$$\lim_{x_i \to +\infty} M_i(x_i) \, e^{|t_i||x_i|} = 0 \qquad for \; |t_i| < \alpha_i, \qquad (i = 1, 2),$$

(b) *there exists a d.f.* $F(x_1, x_2)$, *such that* $\lim_{n \to \infty} F_n(x_1, x_2) = F(x_1, x_2)$ *at each point of continuity of* $F(x_1, x_2)$,

(c) *the m.g.f. of* $F(x_1, x_2)$ *exists for* $|t_i| < \alpha_i$ *and is identically equal to* $\varphi(t_1, t_2)$ *for* $|t_i| < \alpha_i$, $(i = 1, 2)$,

(d) $\lim_{n \to \infty} \varphi_n(t_1, t_2) = \varphi(t_1, t_2)$ *uniformly for* $|t_i| \leq \beta_i < \alpha_i$, $(i = 1, 2)$.

To prove (a), it is sufficient to notice that $\varphi_{1n}(t_1) = \varphi_n(t_1, 0)$ and $\varphi_{2n}(t_2) = \varphi_n(0, t_2)$. Consequently we have

$$\lim_{n \to \infty} \varphi_{1n}(t_1) = \varphi(t_1, 0), \quad \lim_{n \to \infty} \varphi_{2n}(t_2) = \varphi(0, t_2), \, |t_i| < \alpha_i, \quad (i = 1, 2).$$

Therefore (a) follows immediately from Theorem 2.

To prove (b), we may notice that according to the Helly principle of selection applied to the sequence $\{F_n(x_1, x_2)\}$, there exists a subsequence $\{F_{n_k}(x_1, x_2)\}$, selected from the sequence $\{F_n(x_1, x_2)\}$ which is convergent to some function $F(x_1, x_2)$ continuous on the right and with non-negative second difference. But $F(x_1, x_2)$ must be a d.f. according to the Theorem 5, since the relation (6) is satisfied by the sequence $\{F_{n_k}(x_1, x_2)\}$. Moreover, the limit of the sequence $\{\varphi_{n_k}(t_1, t_2)\}$, namely $\varphi(t_1, t_2)$, when considered in a sufficiently small neighborhood of the origin, is the m.g.f. of $F(x_1, x_2)$. Since the d.f. is uniquely determined by its m.g.f., it follows that every convergent subsequence of $\{F_n(x_1, x_2)\}$ converges to the same limit $F(x_1, x_2)$ at each point of continuity of $F(x_1, x_2)$. This is, however, the same as to say that the sequence $\{F_n(x_1, x_2)\}$ itself converges to $F(x_1, x_2)$ at each point of continuity of $F(x_1, x_2)$.

To prove (c), we have to show that the m.g.f. of $F(x_1, x_2)$, say $\varphi^*(t_1, t_2)$, exists for $|t_i| < \alpha_i$ and is equal to $\varphi(t_1, t_2)$, $|t_i| < \alpha_i$, $(i = 1, 2)$. (We have proved that $\varphi^*(t_1, t_2) = \varphi(t_1, t_2)$ only for sufficiently small values of $|t_1|$ and $|t_2|$). The existence of $\varphi^*(t_1, t_2)$ for $|t_i| < \alpha_i$, $(i = 1, 2)$, can be easily established by the method used by Curtiss [1, p. 433]. Suppose indeed that $\varphi^*(t_1, t_2)$ does not exist at some point (t_1^0, t_2^0), where $|t_i^0| < \alpha_i$, $(i = 1, 2)$. That means that we can find a positive number N such that

(17)
$$\int_{-N}^{N} \int_{-N}^{N} e^{t_1^0 x_1 + t_2^0 x_2} \, dF x_1, x_2) > \varphi(t_1^0, t_2^0)$$

Since $\lim F_n(x_1, x_2) = F(x_1, x_2)$ at all points of continuity of $F(x_1, x_2)$, and since

N can be so chosen that the marginal d.f.'s $F_1(x_1)$ and $F_2(x_2)$ are continuous for $x_1 = x_2 = \pm N$, it follows that

$$(18) \quad \lim_{n \to \infty} \int_{-N}^{N} \int_{-N}^{N} e^{t_1^0 x_1 + t_2^0 x_2} dF_n(x_1, x_2) = \int_{-N}^{N} \int_{-N}^{N} e^{t_1^0 x_1 + t_2^0 x_2} dF(x_1, x_2).$$

The formulas (17) and (18) give $\lim_{n \to \infty} \varphi_n(t_1^0, t_2^0) > \varphi(t_1^0, t_2^0)$, which is impossible because $\lim_{n \to \infty} \varphi_n(t_1, t_2) = \varphi(t_1, t_2)$ for $|t_i| < \alpha_i$, $(i = 1, 2)$.

To prove that $\varphi(t_1, t_2) = \varphi^*(t_1, t_2)$ for $|t_i| < \alpha_i$, $(i = 1, 2)$, let (t_1, t_2) denote a fixed point such that $|t_i| < \alpha_i$, $(i = 1, 2)$. Clearly, $\varphi_n(tt_1, tt_2)$, $(n = 1, 2, \cdots)$, and $\varphi^*(tt_1, tt_2)$, considered as functions of the variable t, are m.g.f.'s provided that $|tt_i| < \alpha_i$, $(i = 1, 2)$. (See first part of proof of Theorem 4). Now, according to Theorem 2, the limit of the sequence $\{\varphi_n(tt_1, tt_2)\}$, namely $\varphi(tt_1, tt_2)$, $|tt_i| < \alpha_i$, $(i = 1, 2)$, is also a m.g.f. Since $\varphi(tt_1, tt_2) = \varphi^*(tt_1, tt_2)$ in a sufficiently small interval containing the point $t = 0$, it follows from the uniqueness property of the m.g.f. in the univariate case that $\varphi(tt_1, tt_2) = \varphi^*(tt_1, tt_2)$ identically for $|tt_i| < \alpha_i$, $(i = 1, 2)$. Putting $t = 1$, we find $\varphi(t_1, t_2) = \varphi^*(t_1, t_2)$, $|t_i| < \alpha_i$, $(i = 1, 2)$. Thus (c) is completely proved.

To prove (d), it is sufficient to notice that the sequence $\{\varphi_n(t_1, t_2)\}$ is uniformly continuous in each two-dimensional interval $|t_i| \leq \beta_i < \alpha_i$, $(i = 1, 2)$, (that is, for any $\epsilon > 0$, there exists a positive number $\delta = \delta(\epsilon)$ such that

$$|\varphi_n(t_1', t_2') - \varphi_n(t_1'', t_2'')| < \epsilon$$

if

$$|t_i' - t_i''| < \delta, |t_i'| \leq \beta_i, \quad |t_i''| \leq \beta_i, (i = 1, 2), \qquad (n = 1, 2, \cdots)).$$

Consequently, the sequence $\{\varphi_n(t_1, t_2)\}$ which is convergent for $|t_i| \leq \beta_i$, must be uniformly convergent if $|t_i| \leq \beta_i$, $(i = 1, 2)$.

REFERENCES

[1] J. H. CURTISS, "A note on the theory of moment generating functions", *Annals of Math. Stat.*, Vol. 13 (1942).

[2] D. V. WIDDER, *The Laplace Transform*, Princeton Univ. Press, 1941.

[3] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tract No. 36, 1937.

A GENERALIZATION OF TSHEBYSHEV'S INEQUALITY TO TWO DIMENSIONS

By Z. W. Birnbaum, J. Raymond, and H. S. Zuckerman

University of Washington

1. Let X_1, X_2, $\cdots$, X_n be independent random variables with expectations $E(X_j) = e_j$ and variances $\sigma^2(X_j) = \sigma_j^2$ for $j = 1, 2, \cdots, n$. The question may be asked: What is the upper bound for the probability $P\left(\sum_{j=1}^{n} \frac{(X_j - e_j)^2}{t_j^2} \geq 1\right)$ that the point $(X_1, X_2, \cdots, X_n)$ does not fall inside of the ellipsoid

$$\sum_{j=1}^{n} \frac{(X_j - e_j)^2}{t_j^2} = 1?$$

For $n = 1$ the answer to this question is given by Tshebyshev's inequality

$$(1.1) \qquad P\left[\frac{(X - E(x))^2}{t^2} \geq 1\right] \leq \frac{\sigma^2(X)}{t^2}$$

which can not be improved without further assumptions. By a trivial generalization of the argument leading to (1.1) one can prove the inequality

$$(1.2) \qquad P\left(\sum_{j=1}^{n} \frac{(X_j - e_j)^2}{t_j^2} \geq 1\right) \leq \sum_{j=1}^{n} \frac{\sigma_j^2}{t_j^2}$$

for any integer n. This inequality, however, can be improved for $n \geq 2$. In particular, for $n = 2$, the following theorem will be proved:

THEOREM 1.1. *Let X and Y be independent random variables, with expectations $E(X) = X_0$, $E(Y) = Y_0$ and variances σ_X^2, σ_Y^2. Then, for any $s > 0$, $t > 0$ such that $\frac{\sigma_X^2}{s^2} \leq \frac{\sigma_Y^2}{t^2}$ we have*

$$(1.3) \qquad P\left[\frac{(X - X_0)^2}{s^2} + \frac{(Y - Y_0)^2}{t^2} \geq 1\right] \leq L(s, t)$$

where

$$(1.4)\quad L(s, t) = \begin{cases} 1 \quad \text{if } \dfrac{\sigma_X^2}{s^2} + \dfrac{\sigma_Y^2}{t^2} \geq 1 \\[2em] \dfrac{\sigma_X^2}{s^2} + \dfrac{\sigma_Y^2}{t^2} - \dfrac{\sigma_X^2}{s^2} \cdot \dfrac{1 - \left(\dfrac{\sigma_X^2}{s^2} + \dfrac{\sigma_Y^2}{t^2}\right)}{1 - \dfrac{\sigma_X^2}{s^2}} \quad \text{if} \\[1em] \qquad \dfrac{\sigma_X^2}{s^2} + \dfrac{\sigma_Y^2}{t^2} \leq 1 \leq \frac{1}{2}\left(\dfrac{\sigma_X^2}{s^2} + \dfrac{2\sigma_Y^2}{t^2} + \sqrt{\dfrac{\sigma_X^4}{s^4} + \dfrac{4\sigma_Y^4}{t^4}}\right) \\[2em] \dfrac{\sigma_X^2}{s^2} + \dfrac{\sigma_Y^2}{t^2} - \dfrac{\sigma_X^2\sigma_Y^2}{s^2 t^2} \text{ if } \frac{1}{2}\left(\dfrac{\sigma_X^2}{s^2} + \dfrac{2\sigma_Y^2}{t^2} + \sqrt{\dfrac{\sigma_X^4}{s^4} + \dfrac{4\sigma_Y^4}{t^4}}\right) \leq 1. \end{cases}$$

For any given σ_X^2, σ_Y^2, $s > 0$, $t > 0$ *such that* $\dfrac{\sigma_X^2}{s^2} \leq \dfrac{\sigma_Y^2}{t^2}$ *there exist independent random variables* X *and* Y *with the variances* σ_X^2, σ_Y^2, *such that the equality sign is true in* (1.3).

This theorem is a special case of the more general statement:

THEOREM 1.2. *Let* W, Z *be independent random variables such that*

$$(1.4) \qquad P(W < 0) = P(Z < 0) = 0,$$

$$(1.5) \qquad E(W) = \lambda, E(Z) = \mu,$$

$$(1.6) \qquad \lambda \leq \mu.$$

Then, for any $t > 0$, *we have*

$$(1.7) \qquad P(W + Z \geq t) \leq M(t)$$

where

$$(1.8) \quad M(t) = \begin{cases} 1 & \text{if } t \leq \lambda + \mu \\[2mm] \dfrac{\lambda + \mu}{t} - \dfrac{\lambda}{t} \cdot \dfrac{t - (\lambda + \mu)}{t - \lambda} = \dfrac{\mu}{t - \lambda} \\[2mm] \qquad\qquad \text{if } \lambda + \mu \leq t \leq \tfrac{1}{2}(\lambda + 2\mu + \sqrt{\lambda^2 + 4\mu^2}) \\[3mm] \dfrac{\lambda + \mu}{t} - \dfrac{\lambda\mu}{t^2} & \text{if } \tfrac{1}{2}(\lambda + 2\mu + \sqrt{\lambda^2 + 4\mu^2}) \leq t. \end{cases}$$

For any given $\lambda > 0$, $\mu > 0$, $\lambda \leq \mu$, *and* $t > 0$, *there exist independent variables* W, Z *such that* (1.4) *and* (1.5) *are fulfilled and that the equality sign is true in* (1.7).

Theorem 1.1 is obtained from Theorem 1.2 by writing

$$W = \frac{(X - X_0)^2}{s^2}, \qquad Z = \frac{(Y - Y_0)^2}{t^2}, \qquad t = 1.$$

2. Before proving Theorem 1.2 we shall derive two lemmas. The first of these lemmas deals with more than one variable. Since its proof for general m does not present any additional difficulties it will be stated and proven for any number $m \geq 1$ of variables, although in the proof of Theorem 1.2 it will be used only for $m = 1$.

LEMMA 1. *Let* U, V_1, V_2, $\cdots$, V_m *be independent discrete random variables with only non-negative possible values, and let* U *have a probability distribution with the possible values* $0 \leq U_1 \leq U_2 \leq \cdots \leq U_n$ *and the probabilities* $P(U_i) = r_i$ *for* $i = 1, 2, \cdots, n$. *We consider any three possible values* U_j, U_k, U_l *of* U *such that*

$$0 \leq U_j \leq U_k \leq U_l,$$

with the corresponding probabilities r_j, r_k, r_l. *Then, for any* $t > 0$, *there exists a random variable* U' *with the same distribution as* U *except that the probabilities* r_j, r_k, r_l *of* U_j, U_k, U_l *are replaced by* r_j', r_k', r_l' *such that*

(2.1) $E(U') = E(U)$

(2.2) *one of r'_j, r'_k, r'_l is zero*

(2.3) $P(U' + V_1 + \cdots + V_m \geq t) \geq P(U + V_1 + \cdots + V_m \geq t)$.

PROOF: let r'_j, r'_k, r'_l be written

(2.4) $r'_j = r_j + \alpha\beta, \; r'_k = r_k - \beta, \; r'_l = r_l + (1 - \alpha)\beta$.

For any α, β we then have

$$r'_j + r'_k + r'_l = r_j + r_k + r_l.$$

Choosing

(2.5) $\alpha = (U_l - U_k)/(U_l - U_j)$

we obtain the equality

$$U_j r'_j + U_k r'_k + U_l r'_l = U_j r_j + U_k r_k + U_l r_l$$

so that (2.1) is true for any β.
We obviously have

$$P\left(U + \sum_{s=1}^{m} V_s \geq t\right) = \sum_{i=1}^{n} P(U = U_i) \cdot P\left(\sum_{s=1}^{m} V_s \geq t - U_i\right)$$

(2.6)

$$= \sum_{i=1}^{n} r_i P\left(\sum_{s=1}^{m} V_s \geq t - U_i\right).$$

The variable U' has the same possible values U_i as the variable U. Writing $P(U' = U_i) = r'_i$, for $i = 1, 2, \cdots, n$, we also have

(2.7) $$P\left(U' + \sum_{s=1}^{m} V_s \geq t\right) = \sum_{i=1}^{n} r'_i P\left(\sum_{s=1}^{m} V_s \geq t - U_i\right).$$

From (2.6), (2.7), and (2.4) we obtain

$$P\left(U' + \sum_{s=1}^{m} V_s \geq t\right) - P\left(U + \sum_{s=1}^{m} V_s \geq t\right)$$

(2.8)

$$= \alpha\beta P\left(\sum_{s=1}^{m} V_s \geq t - U_j\right) - \beta P\left(\sum_{s=1}^{m} V_s \geq t - U_k\right)$$

$$+ (1 - \alpha)\beta P\left(\sum_{s=1}^{m} V_s \geq t - V_l\right).$$

For α determined by (2.5), the right-hand side of (2.8) is of the form $C\beta$, and will be positive if sign β = sign C. If sign C is positive, we choose $\beta = r_k$ and have, from (2.4), $r'_k = 0$, and, from (2.8), the inequality (2.3). If sign C is negative, we set $\beta = \text{Max}\left(-\dfrac{r_i}{\alpha}, -\dfrac{r_l}{1 - \alpha}\right)$ which leads to either $r'_j = 0$ or $r'_l = 0$, and again to (2.3). In both cases we have kept the probabilities r'_j, r'_k, r'_l non-negative as they should be.

LEMMA 2. *Let the discrete random variable U have only the two non-negative values $U_1 < U_2$, with the corresponding probabilities r_1, r_2, and let t be a given number such that*

$$(2.9) \qquad E(U) < t < U_2 .$$

Then there exists a number $\alpha \geq 0$ such that the random variable U' with the possible values

$$(2.91) \qquad U_1' = U_1 + \alpha$$
$$U_2' = t$$

and the corresponding probabilities r_1, r_2, has the properties

$$(2.92) \qquad 0 \leq U_1' \leq U_2'$$
$$(2.93) \qquad E(U') = E(U).$$

PROOF: to have (2.91) and (2.93) it is sufficient to choose

$$\alpha = \frac{r_2(U_2 - t)}{r_1} .$$

Then (2.92) is also fulfilled since, in view of (2.9), we have

$$U_1' = \frac{r_1 U_1 + r_2 U_2 - r_2 t}{r_1} = \frac{E(U) - r_2 t}{r_1} \leq \frac{t - r_2 t}{r_1} = t = U_2',$$

and obviously $\alpha \geq 0$ and hence $U_1' \geq U_1 \geq 0$.

3. Theorem 1 will first be proven under the assumption that W and Z are discrete random variables, each with a finite number of non-negative possible values. By repeatedly applying Lemma 1 with $m = 1$, $U = W$, $V_1 = Z$, we reduce the number of possible values of W which have non-zero probabilities to two, and denote those possible values by $W_1 \leq W_2$, and their probabilities by p_1 and $p_2 = 1 - p_1$. Then, applying Lemma 1 to the case $m = 1$, $U = Z$, $V_1 = W$, we similarly reduce the possible values of Z to the two non-negative values $Z_1 \leq Z_2$, and denote the corresponding probabilities by q_1 and $q_2 = 1 - q_1$. Throughout all these steps the expectations $E(W) = \lambda$ and $E(Z) = \mu$ remain unchanged, and $P(W + Z \geq t)$ is not decreased.

For $t \leq \lambda + \mu$, inequality (1.3) is obviously true, and equality is attained for W having the only possible value λ with probability 1 and Z having the only possible value μ with probability 1.

For the remainder of the proof we assume $t > \lambda + \mu$. We then have

$$t > \lambda + \mu \geq \lambda + Z_1 \geq W_1 + Z_1 .$$

If $W_2 > t$, we may replace it by $W_2 = t$ according to Lemma 2. Similarly, if $Z_2 > t$, we may replace it by $Z_2 = t$. The probability $P(W + Z \geq t)$ is not decreased in this process. We may thus assume, without loss of generality, that

$$W_2 \leq t, \qquad Z_2 \leq t.$$

The joint distribution of (W, Z) has now the possible values represented by the four points (W_1, Z_1), (W_1, Z_2), (W_2, Z_1), (W_2, Z_2). The coordinates of these four points and their probabilities fulfill the following conditions

$$(3.1) \qquad 0 \leq W_1 \leq \lambda \leq W_2 \leq t; \qquad 0 \leq Z_1 \leq \mu \leq Z_2 \leq t$$

$$(3.2) \qquad p_1 + p_2 = q_1 + q_2 = 1$$

$$(3.3) \qquad p_1 W_1 + p_2 W_2 = \lambda, \qquad q_1 Z_1 + q_2 Z_2 = \mu.$$

In view of (3.1), the point (W_1, Z_1) always lies below the line $W + Z = t$. The other points may or may not lie below that line. Accordingly, we distinguish the cases listed in Table I. These clearly include all possible cases since (W_2, Z_2) can not be below the line $W + Z = t$ without all the other points being below that line.

In case V we have $P(W + Z \geq t) = 0$.

For the discussion of the remaining cases we note the following relationships which follow from (3.2) and (3.3).

<div align="center">TABLE I</div>

Case	Points below line $W + Z = t$	Points not below line $W + Z = t$
I	(W_1, Z_1)	(W_2, Z_1), (W_1, Z_2), (W_2, Z_2)
II	(W_1, Z_1), (W_2, Z_1)	(W_1, Z_2), (W_2, Z_2)
III	(W_1, Z_1), (W_1, Z_2)	(W_2, Z_1), (W_2, Z_2)
IV	(W_1, Z_1), (W_2, Z_1), (W_1, Z_2)	(W_2, Z_2)
V	(W_1, Z_1), (W_2, Z_1), (W_1, Z_2), (W_2, Z_2)	none

$$p_1 = \frac{W_2 - \lambda}{W_2 - W_1}, \qquad p_2 = \frac{\lambda - W_1}{W_2 - W_1},$$

$$q_1 = \frac{Z_2 - \mu}{Z_2 - Z_1}, \qquad q_2 = \frac{\mu - Z_1}{Z_2 - Z_1}.$$

In case I we have

$$(3.41) \qquad W_1 + Z_1 < t, \quad W_2 + Z_1 \geq t, \quad W_1 + Z_2 \geq t, \quad W_2 + Z_2 \geq t,$$

$$P = P(W + Z \geq t) = p_2 q_1 + p_1 q_2 + p_2 q_2 = 1 - p_1 q_1$$

$$1 - \frac{W_2 - \lambda}{W_2 - W_1} \cdot \frac{Z_2 - \mu}{Z_2 - Z_1}.$$

Since P is a decreasing function of W_1 and Z_1, we replace W_1 and Z_1 by the smallest values compatible with (3.41), namely $W_1 = t - Z_2$, $Z_1 = t - W_2$, and obtain

$$P \leq 1 - \frac{(W_2 - \lambda)(Z_2 - \mu)}{(W_2 + Z_2 - t)^2} = R(W_2, Z_2).$$

For fixed Z_2, $R(W_2, Z_2)$ has a minimum at $W_2 = Z_2 + 2\lambda - t$ and no other extremum, hence it assumes its maximum at one or both of the end-points of the interval for W_2 which, by (3.1) and (3.41), is

$$t - Z_1 \leq W_2 \leq t.$$

In view of (3.1) we also have $t - \mu \leq t - Z_1$, and hence

$$P \leq \text{Max } [R(t - \mu, Z_2), \quad R(t, Z_2)].$$

We find

$$R(t - \mu, Z_2) = 1 - \frac{t - \mu - \lambda}{Z_2 - \mu} \leq 1 - \frac{t - \mu - \lambda}{t - \mu} = \frac{\lambda}{t - \mu}$$

and

$$R(t, Z_2) = 1 - \frac{(t - \lambda)(Z_2 - \mu)}{Z_2^2} = R^{(1)}(Z_2).$$

This last expression has a minimum for $Z_2 = 2$ and no other extremum, hence it assumes its maximum at the ends of the interval for Z_2 which, by (3.41) and (3.1), is

$$t - W_1 \leq Z_2 \leq t.$$

From (3.1) we also have $t - \lambda \leq t - W_1$; and thus

$$R(t, W_2) \leq \text{Max } [R^{(1)}(t - \lambda), \quad R^{(1)}(t)] = \text{Max } \left[\frac{\mu}{t - \lambda}, \quad \frac{\lambda + \mu}{t} - \frac{\lambda\mu}{t^2} \right].$$

Finally, we obtain

$$P \leq \text{Max } \left[\frac{\lambda}{t - \mu}, \quad \frac{\mu}{t - \lambda}, \quad \frac{\lambda + \mu}{t} - \frac{\lambda\mu}{t_2} \right].$$

Each of the values $P = \dfrac{\lambda}{t - \mu}, \dfrac{\mu}{t - \lambda}, \dfrac{\lambda + \mu}{t} - \dfrac{\lambda\mu}{t^2}$ can be attained in case **I**, as is shown by the probability distributions

$$W_1 = 0, \quad W_2 = t - \mu, \quad Z_1 = \mu, \quad Z_2 = t,$$

(3.42)
$$p_1 = 1 - \frac{\lambda}{t - \mu}, \quad p_2 = \frac{\lambda}{t - \mu}, \quad q_1 = 1, \quad q_2 = 0;$$

$$W_1 = \lambda, \quad W_2 = t, \quad Z_1 = 0, \quad Z_2 = t - \lambda,$$

(3.43)
$$p_1 = 1, \quad p_2 = 0, \quad q_1 = 1 - \frac{\mu}{t - \lambda}, \quad q_2 = \frac{\mu}{t - \lambda};$$

$$\cdot W_1 = 0, \quad W_2 = t, \quad Z_1 = 0, \quad Z_2 = t,$$

(3.44)
$$p_1 = 1 - \frac{\lambda}{t}, \quad p_2 = \frac{\lambda}{t}, \quad q_1 = 1 - \frac{\mu}{t}, \quad q_2 = \frac{\mu}{t}.$$

In case II we have

$$W_1 + Z_1 < t, \qquad W_2 + Z_1 < t, \qquad W_1 + Z_2 \geq t, \qquad W_2 + Z_2 \geq t,$$

(3.51)

$$P = P(W + Z \geq t) = p_1 q_2 + p_2 q_2 = q_2 = \frac{\mu - Z_1}{Z_2 - Z_1}.$$

This is a decreasing function of Z_1 as well as of Z_2 and hence takes its maximum for the smallest values of Z_1 and Z_2 compatible with (3.1) and (3.5), that is for $Z_1 = 0$, $Z_2 = t - \lambda$. We thus obtain

$$P \leq \frac{\mu}{t - \lambda}.$$

This upper bound can be attained in case II, as may be seen from the distribution

$$W_1 = \lambda, \qquad W_2 = \lambda, \qquad Z_1 = 0, \qquad Z_2 = t - \lambda,$$

(3.52)

$$p_1 = \tfrac{1}{2}, \qquad p_2 = \tfrac{1}{2}, \qquad q_1 = 1 - \frac{\mu}{t - \lambda}, \qquad q_2 = \frac{\mu}{t - \lambda}.$$

Case III is symmetrical with case II and leads to the inequality

$$P \leq \frac{\lambda}{t - \mu}.$$

In case IV we have

$$W_1 + Z_1 < t, \qquad W_2 + Z_1 < t, \qquad W_1 + Z_2 < t, \qquad W_2 + Z_2 \geq t,$$

(3.61)

$$P = P(W + Z \geq t) = p_2 q_2 = \frac{(\lambda - W_1)(\mu - Z_1)}{(W_2 - W_1)(Z_2 - Z_1)}.$$

The right hand side is a decreasing function of each of the variables W_1, W_2, Z_1, Z_2, and hence is increased by chosing for these variables the smallest values compatible with (3.61), i.e.

(3.62) $$W_1 = Z_1 = 0, \qquad W_2 + Z_2 = t$$

for which we obtain

$$P \leq \frac{\lambda}{W_2} \frac{\mu}{t - W_2} = R^{(2)}(W_2).$$

Since $R^{(2)}(W_2)$ has a minimum at $W_2 = \frac{t}{2}$ and no other extremum, it attains its largest value at one of the end points of the interval for W_2 which, by (3.1), (3.61) and (3.62), is

$$\lambda \leq W_2 \leq t - \mu.$$

This leads to

$$P \leq \mathrm{Max}\,[R^{(2)}(\lambda), \qquad R^{(2)}(t - \mu)] = \mathrm{Max}\left[\frac{\mu}{t - \lambda}, \qquad \frac{\lambda}{t - \mu}\right].$$

The upper bounds $\dfrac{\mu}{t-\lambda}$, $\dfrac{\lambda}{t-\mu}$, respectively, are attained in case IV for the probability distribution

$$W_1 = 0, \qquad W_2 = \lambda, \qquad Z_1 = 0, \qquad Z_2 = t - \lambda,$$

(3.63)
$$p_1 = 0, \qquad p_2 = 1, \qquad q_1 = 1 - \frac{\mu}{t-\lambda}, \qquad q_2 = \frac{\mu}{t-\lambda}$$

and

$$W_1 = 0, \qquad W_2 = t - \mu, \qquad Z_1 = 0, \qquad Z_2 = \mu,$$

$$p_1 = 1 - \frac{\lambda}{t-\mu}, \qquad p_2 = \frac{\lambda}{t-\mu}, \qquad q_1 = 0, \qquad q_2 = 1.$$

From the preceding discussion we conclude that $P = P(W + Z \geq t)$ always fulfills the inequality

$$P \leq \text{Max}\left[\frac{\lambda}{t-\mu}, \quad \frac{\mu}{t-\lambda}, \quad \frac{\lambda+\mu}{t} - \frac{\lambda\mu}{t^2}\right] = U(t)$$

for $t \geq \lambda + \mu$. Since we have assumed $\lambda \leq \mu$, we have $\dfrac{\lambda}{t-\mu} \leq \dfrac{\mu}{t-\lambda}$ for $t \geq \lambda + \mu$, and therefore

$$U(t) = \text{Max}\left[\frac{\mu}{t-\lambda}, \quad \frac{\lambda+\mu}{t} - \frac{\lambda\mu}{t^2}\right] \text{ for } t \geq \lambda + \mu.$$

It is easily verified that

$$\frac{\mu}{t-\lambda} \leq \frac{\lambda+\mu}{t} - \frac{\lambda\mu}{t^2} \text{ for } \lambda + \mu \leq t \leq \tfrac{1}{2}\left(\lambda + 2\mu + \sqrt{\lambda^2 + 4\mu^2}\right)$$

and

$$\frac{\mu}{t-\lambda} \leq \frac{\lambda+\mu}{t} - \frac{\lambda\mu}{t^2} \text{ for } \tfrac{1}{2}\left(\lambda + 2\mu + \sqrt{\lambda^2 + 4\mu^2}\right) \leq t$$

so that we have $U(t) = M(t)$ as defined in (1.8). For given λ, μ and any $t \geq \lambda + \mu$, the equality $P = \dfrac{\mu}{t-\lambda}$ is fulfilled for the distributions (3.43), (3.52) and (3.63), while the equality $P = \dfrac{\lambda+\mu}{t} - \dfrac{\lambda\mu}{t^2}$ is true for the distribution (3.44).

This completes the proof of Theorem 1.2 for discrete random variables. If W and Z are independent random variables with the cumulative probability functions $P(W \leq w) = F(w)$ and $P(Z \leq z) = G(z)$, then each of these cumulative probability functions can be uniformly approximated by a step function with a finite number of steps, that is by the cumulative probability function of a discrete random variable with a finite number of possible values. Since for such variables Theorem 1.2 is proven, it also is true for the general random variables W and Z.

4. An attempt to extend the method used in proving Theorem 1.2 to more than two variables leads to arguments of a prohibitive length. It is possible, however, to obtain corollaries of Theorems 1.1 and 1.2 which lead to an improvement of inequality (1.2) for n variables.

COROLLARY 2.1   *Let $X_1, X_2, \cdots, X_n$ be independent random variables with expectations $E(X_j) = e_j$ and variances $\sigma^2(X_j) = \sigma_j^2$. Then, for any $t_j \geq 0$, $j = 1, 2, \cdots, n$, and any m such that*

$$\Sigma_1 = \sum_{j=1}^{m} \frac{\sigma_j^2}{t_j^2} \leq \sum_{j=m+1}^{n} \frac{\sigma_j^2}{t_j^2} = \Sigma_2,$$

we have the inequality

$$P\left(\sum_{j=1}^{n} \frac{(X_j - e_j)^2}{t_j^2} \geq 1\right) \leqq \begin{cases} 1 & \text{if} \quad 1 \leq \sum_{j=1}^{n} \frac{\sigma_j^2}{t_j^2} = \Sigma_1 + \Sigma_2 \\[2mm] \sum_{j=1}^{n} \frac{\sigma_j^2}{t_j^2} - \Sigma_1 \cdot \dfrac{t - (\Sigma_1 + \Sigma_2)}{t - \Sigma_1} \\ \qquad \text{if} \quad \Sigma_1 + \Sigma_2 \leq t \leq \tfrac{1}{2}\left[\Sigma_1 + 2\Sigma_2 + \sqrt{\Sigma_1^2 + 4\Sigma_2^2}\right] \\[2mm] \sum_{j=1}^{n} \frac{\sigma_j^2}{t_j^2} - \Sigma_1 \cdot \Sigma_2 & \text{if} \quad \tfrac{1}{2}\left[\Sigma_1 + 2\Sigma_2 + \sqrt{\Sigma_1^2 + 4\Sigma_2^2}\right] \leq t. \end{cases}$$

This corollary is a special case of the following corollary to Theorem 1.2

COROLLARY 2.2.   *Let $W_1, W_2, \cdots, W_n$ be independent random variables such that $P(W_j < 0) = 0$ for $j = 1, 2, \cdots, n$, and let m be any integer such that*

$$\sum_{j=1}^{m} E(W_j) = \lambda, \qquad \sum_{j=m+1}^{n} E(W_j) = \mu, \qquad \lambda \leq \mu.$$

Then, for any $t \geq 0$, we have

$$P\left(\sum_{j=1}^{n} W_j \geq t\right) \leq M(t)$$

where $M(t)$ is defined by (1.8).

This corollary follows immediately from Theorem 1.2 by writing

$$W = \sum_{j=1}^{m} W_j, \qquad Z = \sum_{j=m+1}^{n} W_j.$$

To obtain Corollary 2.1, one only has to write in Corollary 2.2

$$W_j = \frac{(X_j - e_j)^2}{t_j^2}.$$

If some additional assumptions are made on the expectations $E(W_j)$ or on the

variances σ_j^2, the upper bounds in Corollaries 2.1 and 2.2 may be minimized by proper choice of m or of the t_j. For example, if all the variances are equal

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$$

and n is even, one obtains the inequality

$$P\left[\sum_{j=1}^{n}(X_j - e_j)^2 \geq t^2\right] \leq \begin{cases} 1 & \text{if} \quad t^2 \leq n\sigma^2 \\[2mm] 1 - \dfrac{t^2 - n\sigma^2}{t^2 - \dfrac{n}{2}\sigma^2} & \text{if } n\sigma^2 \leq t^2 \leq \dfrac{3 + \sqrt{5}}{4}\, n\sigma^2 \\[2mm] \dfrac{n\sigma^2}{t^2}\left(1 - \dfrac{1}{4}\dfrac{n\sigma^2}{t^2}\right) & \text{if} \quad \dfrac{3 + \sqrt{5}}{4}\, n\sigma^2 \leq t^2 \, . \end{cases}$$

DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT IN A CIRCULARLY CORRELATED UNIVERSE[1]

By R. B. Leipnik

Cowles Commission for Research in Economics

1. Summary. It is desired to find an approximate distribution of simple form for the statistic $\bar{r} = \dfrac{x_1 x_2 + \cdots + x_T x_1}{x_1^2 + \cdots + x_T^2}$ ($\bar{r}$ is an estimate of the serial correlation coefficient ρ in a circular universe) in the case that $\rho \neq 0$ in the universe. Such a distribution is obtained by smoothing the joint characteristic function of the numerator and denominator of the expression for $\bar{r}$. The first two moments are calculated; from these $\bar{r}$ is seen to be a consistent estimate of ρ. A graph of this distribution for sample size $T = 20$ and various values of ρ is given.

In addition, an approximate distribution for $p = x_1^2 + \cdots + x_T^2$ is derived which reduces to the exact $(\chi^2$-) distribution if $\rho = 0$. From a formula which yields all moments, it is concluded that, at least up to the degree of approximation attained, p/T is an unbiased and consistent extimate of σ^2.

2. Several writers have investigated the temporally homogeneous stochastic process defined by

$$(1) \qquad x_t - \rho x_{t-1} = z_t, \qquad t = 1, 2, \cdots, T, \; |\rho| < 1,$$

where the z_t are unobservable disturbances, normally and independently distributed with mean zero and variance σ^2, the x_t are observed variates, and the "first observation" x_0 has a normal distribution with mean zero and such a variance σ_x^2 that all later observations have the same variance. Thus we have

$$(2) \qquad \sigma_x^2 = \frac{\sigma^2}{1 - \rho^2}$$

and the joint distribution of a sample of $T + 1$ successive values is

$$
(3) \qquad g(x_0, x_1, \cdots, x_T) = \frac{(1 - \rho^2)^{\frac{1}{2}}}{(2\pi\sigma^2)^{T/2+1/2}} \cdot \exp\left[-\frac{1}{2\sigma^2} \{x_0^2 + x_T^2 \right.
$$
$$
\left. - 2\rho(x_0 x_1 + \cdots + x_{T-1} x_T) + (1 + \rho^2)(x_1^2 + \cdots + x_{T-1}^2)\} \right].
$$

Koopmans ([1], formula 96), by smoothing characteristic values, has obtained an approximation to the distribution of the serial correlation coefficient r for the case $\rho = 0$, where

$$(4) \qquad r = \frac{x_0 x_1 + \cdots + x_{T-1} x_T}{x_0^2 + \cdots + x_T^2}.$$

This result is expressed in the form of a definite integral whose evaluation has not so far been effected.

By considering the related circular stochastic process, where x_0 is defined to be the same observation as x_T, great simplification is obtained. Here the joint distribution of $x_1, x_2, \cdots, x_T$ is

$$f(x_1, x_2, \cdots, x_T) = \frac{\lambda(\rho)}{(2\pi\sigma^2)^{T/2}} \exp\left[- \frac{1}{2\sigma^2(1 - \rho^2)} \right.$$

(5)
$$\left. \{(1 + \rho^2)(x_1^2 + \cdots + x_T^2) - 2\rho(x_1 x_2 + \cdots + x_T x_1)\} \right]$$

$$\lambda(\rho) = \frac{1 - \rho^T}{(1 - \rho^2)^{T/2}}.$$

By smoothing characteristic values, Koopmans ([1], formula 92) found a definite integral and Dixon ([2], 3.22) an explicit expression for an approximate distribution of the circular serial correlation coefficient $\bar{r}$, for the case $\rho = 0$, where

(6)
$$\bar{r} = \frac{x_1 x_2 + \cdots + x_T x_1}{x_1^2 + \cdots + x_T^2}.$$

Dixon's distribution $\bar{R}_0(\bar{r})$ has the simple form

(7)
$$\bar{R}_0(\bar{r}) = \frac{\Gamma\left(\dfrac{T}{2} + 1\right)}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} (1 - \bar{r}^2)^{T/2-\frac{3}{2}}.$$

Rubin [3] proved these results to be equivalent. On the other hand, R. L. Anderson [4] obtained the exact distribution of $\bar{r}$ in the case $\rho = 0$. Madow [5] extended this result to the case $\rho \neq 0$, using a property of sufficient statistics also noted by Koopmans ([1], p. 17) in connection with the non-circular problem.

It would, however, be difficult to find percentile points or moments from Madow's exact distribution. An approximate distribution of $\bar{r}$ for $\rho \neq 0$, together with its moments, analogous to Dixon-Koopmans' for $\rho = 0$, should therefore be of interest. The purpose of this paper is to obtain such a distribution from the circular universe (5). The statistic $\bar{r}$ is shown to be a consistent estimate of ρ within the limits imposed by the approximation. In addition, an approximate distribution for $p = x_1^2 + \cdots + x_T^2$ in the case $\rho \neq 0$ (which reduces to the exact chi-squared distribution when $\rho = 0$) is derived, together with all of its moments.

3. We begin by asking about an approximate joint distribution of p and $\bar{q}$ defined by

(8)
$$p = x_1^2 + \cdots + x_T^2$$
$$\bar{q} = x_1 x_2 + \cdots + x_T x_1.$$

Defining $\phi(u, v)$ as the expectation of $\exp[i(up + v\bar{q})]$, we have

$$(9) \quad \phi(u, v) = \frac{\lambda(\rho)}{(2\pi\sigma^2)^{T/2}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2\sigma^2}\left\{\left(\frac{1 + \rho^2}{1 - \rho^2} - 2i\sigma^2 u\right)p \right.\right.$$
$$\left.\left. - 2\left(\frac{\rho}{1 - \rho^2} + i\sigma^2 v\right)\bar{q}\right\}\right] dx_1 \cdots dx_T.$$

On integration, we find

$$(10) \quad \phi(u, v) = \lambda(\rho)[A(u, v)]^{-\frac{1}{2}}$$

where $A(u, v)$ is the determinant of the matrix associated with the quadratic form within the curly brackets in (9). $A(u, v)$ is a circulant; its value as determined from the circulant formula ([2], p. 123) is

$$(11) \quad A(u, v) = \prod_{t=1}^{T}\left(y - 2z\cos\frac{2\pi t}{T}\right)$$

where y and z are defined by

$$(12) \quad \begin{aligned} y &= \frac{1 + \rho^2}{1 - \rho^2} - 2i\sigma^2 u \\ z &= \frac{\rho}{1 - \rho^2} + i\sigma^2 v. \end{aligned}$$

To get an approximation $\tilde{A}(u, v)$ to $A(u, v)$ we smooth $\log A(u, v)$ by Koopmans' method. We have

$$(13) \quad \log A(u, v) = \sum_{t=1}^{T}\log\left(y - 2z\cos\frac{2\pi t}{T}\right).$$

We define $\tilde{A}(u, v)$ through

$$(14) \quad \log \tilde{A}(u, v) = \int_{0}^{T}\log\left(y - 2z\cos\frac{2\pi t}{T}\right)dt$$

in which the summation in (13) is replaced by integration. The integral in (14) is easily evaluated ([6], p. 65) giving

$$(15) \quad \tilde{A}(u, v) = \left(\frac{y + \sqrt{y^2 - 4z^2}}{2}\right)^{T}.$$

Incidentally, had we used $\bar{q}_L = x_1 x_{L+1} + \cdots + x_T x_{T+L}$ in place of $\bar{q}_1 = \bar{q}$ in (9), we would have obtained the same expression (15) for $\tilde{A}(u, v)$.

Setting $\tilde{\phi}(u, v) = \tilde{\lambda}(\rho)[\tilde{A}(u, v)]^{-\frac{1}{2}}$ we may determine $\tilde{\lambda}(\rho)$ by the requirement $\tilde{\phi}(0, 0) = 1$. A simple calculation yields the result $\tilde{\lambda}(\rho) = (1 - \rho^2)^{-(T/2)}$. (Note that $\frac{\lambda(\rho)}{\tilde{\lambda}(\rho)} = 1 - \rho^T$ is close to 1 for large values of T). Our result for $\tilde{\phi}(u, v)$ appears as

$$(16) \quad \tilde{\phi}(u, v) = \tilde{\lambda}(\rho)\left(\frac{y + \sqrt{y^2 - 4z^2}}{2}\right)^{-(T/2)}$$

The approximate joint distribution of p and $\bar{q}$ may be written as the double Fourier integral

$$(17) \quad \bar{D}(p, \bar{q}) = \frac{\bar{\lambda}(\rho)}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left[-i(up + v\bar{q})\right] \left(\frac{y + \sqrt{y^2 - 4z^2}}{2}\right)^{-T/2} du \, dv$$

which we evaluate ([7], 576.3, 914.3) by changing integration variables from u, v to y, z and integrating out y and z successively. We obtain finally

$$\bar{D}(p, \bar{q}) = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} p^{-(T/2)-1}(p^2 - \bar{q}^2)^{T/2-\frac{1}{2}}$$

$$(18) \qquad\qquad\qquad \cdot \exp\left[-\frac{1}{2\sigma^2(1 - \rho^2)}\{(1 + \rho^2)p - 2\rho\bar{q}\}\right].$$

Changing variables from p, $\bar{q} = p\bar{r}$ to p, $\bar{r}$, we obtain for $\bar{F}(p, \bar{r})$, the approximate joint distribution of p and $\bar{r}$, the expression

$$\bar{F}(p, \bar{r}) = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-(T/2)}}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} p^{T/2-1}(1 - \bar{r}^2)^{T/2-\frac{1}{2}}$$

$$(19) \qquad\qquad\qquad \cdot \exp\left[-\frac{p}{2\sigma^2(1 - \rho^2)}\{1 + \rho^2 - 2\rho\bar{r}\}\right].$$

We could also have derived (19), following Madow, by noting that for $\rho = 0$, p and $\bar{r}$ are independently distributed, p having the chi-squared distribution and $\bar{r}$ having approximately the Dixon distribution (7), and that p and $\bar{r}$ are sufficient statistics for the estimation of ρ and σ^2.

4. The approximate marginal distribution $\bar{R}_\rho(\bar{r})$ of $\bar{r}$ is obtained by an easy integration from (19)

$$\bar{R}_\rho(\bar{r}) = \int_0^\infty \bar{F}(p, \bar{r})dp = \frac{T}{2} \frac{[2\sigma^2(1 - \rho^2)]^{-(T/2)}}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} (1 - \bar{r}^2)^{T/2-\frac{1}{2}}$$

$$(20) \qquad\qquad \cdot \int_0^\infty p^{T/2-1} \exp\left[-\frac{p}{2\sigma^2(1 - \rho^2)}\{1 + \rho^2 - 2\rho\bar{r}\}\right]dp \, ,$$

$$\bar{R}_\rho(\bar{r}) = \frac{\Gamma\left(\dfrac{T}{2} + 1\right)}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} (1 - \bar{r}^2)^{T/2-\frac{1}{2}}(1 + \rho^2 - 2\rho\bar{r})^{-T/2}.$$

Our notation is consistent since $\bar{R}_\rho(\bar{r})$ indeed reduces to the Dixon distribution for $\rho = 0$. $\bar{R}_\rho(\bar{r})$ has a maximum when

$$\bar{r} = \bar{r}_{max} = \frac{1}{2\rho(T - 2)}\{(1 + \rho^2)(T - 1) - \sqrt{T(T - 2)(1 - \rho^2)^2 + (1 + \rho^2)^2}\}.$$

A little manipulation shows that $1 > |\bar{r}_{max}| > |\rho|$ and that $\bar{r}_{max} = \rho$ asymptotically. A graph (Fig. 1) of $\check{R}_\rho(\bar{r})$ for $T = 20$, $\rho = 0$, .2, .5, .7, .9 is appended from which it is seen that for $|\rho|$ near 1, the distribution becomes highly concentrated about $\bar{r}_{max}$. On differentiating $\check{R}_\rho(\bar{r})$ with respect to ρ and eliminating, the envelope of the $\check{R}_\rho(\bar{r})$ is seen to be

$$\frac{\Gamma\left(\dfrac{T}{2} + 1\right)}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)} (1 - \bar{r}^2)^{-\frac{1}{2}}.$$
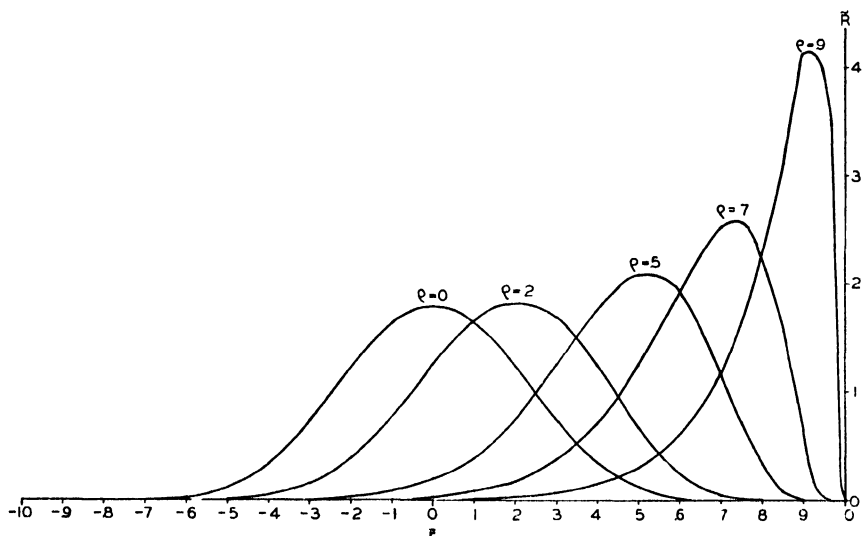
Fig. 1. Graph of the Distribution of the Serial Correlation Coefficient in a Circular Universe, for $T = 20$

5. Before evaluating the moments of $\check{R}_\rho(\bar{r})$ we will pause to obtain the approximate marginal distribution $\check{P}_\rho(p)$ of p, and its moments. We write

$$(21) \quad \check{P}_\rho(p) = \int_{-1}^{+1} \check{F}(p, \bar{r})\, d\bar{r} = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma(\tfrac{1}{2})\Gamma\left(\dfrac{T}{2} + \dfrac{1}{2}\right)}$$

$$\cdot p^{T/2-1} \exp\left[-\frac{p}{2\sigma^2}\left(\frac{1 + \rho^2}{1 - \rho^2}\right)\right] \cdot \int_{-1}^{+1} (1 - \bar{r}^2)^{T/2-\frac{1}{2}} \exp\left[\frac{\rho p \bar{r}}{\sigma^2(1 - \rho^2)}\right] d\bar{r}.$$

If we define $I_\nu(z)$, the Bessel function of order ν and purely imaginary argument by

$$(22) \qquad I_\nu(z) = \sum_{n=0}^{\infty} \frac{\left(\dfrac{z}{2}\right)^{\nu+2n}}{n!\Gamma(\nu + n + 1)},$$

we obtain ([8], p. 79), if $\rho \neq 0$

$$(23) \qquad \tilde{P}_\rho(p) = \frac{T}{2}\, \rho^{-T/2} p^{-1} \exp\left[-\frac{p}{2\sigma^2}\left(\frac{1 + \rho^2}{1 - \rho^2}\right)\right] I_{T/2}\left(\frac{\rho p}{\sigma^2(1 - \rho^2)}\right),$$

and if $\rho = 0$

$$(24) \qquad \tilde{P}_0(p) = \frac{(2\sigma^2)^{-T/2}}{\Gamma\left(\dfrac{T}{2}\right)}\, p^{T/2-1} \exp\left[-\frac{p}{2\sigma^2}\right],$$

on performing the integration indicated in (21). $\tilde{P}_0(p)$ coincides with the exact distribution $P_0(p)$. An expression covering all moments of $\tilde{P}_\rho(p)$ is obtained from (16) by setting $v = 0$, differentiating, and setting $u = 0$. We have

$$(25) \qquad \tilde{\phi}(u, 0) = \tilde{\lambda}(\rho)\left(\frac{y + \sqrt{y^2 - \left[\dfrac{2\rho}{1 - \rho^2}\right]^2}}{2}\right)^{-T/2}$$

hence

$$(26) \qquad \tilde{E}[p^k] = i^{-k} \frac{d^k}{du^k}\, \tilde{\phi}(u, 0)\Bigg|_{u=0} = (-2\sigma^2)^k(1 - \rho^2)^{-T/2}$$

$$\cdot \frac{d^k}{dy^k}\left(\frac{y + \sqrt{y^2 - \left[\dfrac{2\rho}{1 - \rho^2}\right]^2}}{2}\right)^{-T/2}\Bigg|_{y=(1+\rho^2)/(1-\rho^2)}$$

From (26), we readily find

$$(27) \qquad \tilde{E}[p] = T\sigma^2, \qquad \tilde{E}\left[\frac{p}{T}\right] = \sigma^2$$

$$(28) \qquad \tilde{E}[p^2] = (T\sigma^2)^2 + 2T\sigma^4\left(\frac{1 + \rho^2}{1 - \rho^2}\right)$$

$$\tilde{\sigma}_p^2 = 2T\sigma^4\left(\frac{1 + \rho^2}{1 - \rho^2}\right), \qquad \tilde{\sigma}_{p/T}^2 = \frac{2\sigma^4}{T}\left(\frac{1 + \rho^2}{1 - \rho^2}\right).$$

Thus the unbiased character of p/T as an estimate of σ^2 is reflected in the approximate distribution, while (28), which shows that $\lim_{T \to \infty} \tilde{\sigma}_{p/T}^2 = 0$, indicates that consistency is also reflected.

6. We now calculate the moments of $\tilde{R}_\rho(\bar{r})$. Interchanging the order of integration in the expression for $\tilde{E}[\bar{r}^k]$ is justified by the uniform convergence, so we have

$$\bar{E}[\bar{r}^k] = \int_{-1}^{+1} \bar{r}^k \left[\int_0^\infty \bar{F}[p, \bar{r}] \, dp \right] d\bar{r} = \int_0^\infty \left[\int_{-1}^{+1} \bar{r}^k \bar{F}(p, \bar{r}) \, d\bar{r} \right] dp$$

$$(29) \qquad = \frac{T}{2} \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} \int_0^\infty p^{T/2 - 1}$$

$$\cdot \exp\left[-\frac{p}{2\sigma^2} \left(\frac{1 + \rho^2}{1 - \rho^2} \right) \right] \left\{ \int_{-1}^{+1} \bar{r}^k (1 - \bar{r}^2)^{T/2 - 1/2} \exp(m\bar{r}) \, d\bar{r} \right\} dp$$

where m is defined by

$$(30) \qquad m = \frac{\rho p}{\sigma^2(1 - \rho^2)}.$$

Defining $G(m)$ by

$$(31) \qquad G(m) = \int_{-1}^{+1} (1 - \bar{r}^2)^{T/2 - 1/2} \exp(m\bar{r}) \, d\bar{r}$$

we have ([8], p. 79)

$$(32) \qquad G(m) = \left(\frac{m}{2} \right)^{-T/2} \frac{I_{T/2}(m)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)}.$$

Differentiating each side of (32) k times, we find by (31) and (32)

$$\frac{d^k}{dm^k} G(m) = \int_{-1}^{+1} \bar{r}^k (1 - \bar{r}^2)^{T/2 - 1/2} \exp(m\bar{r}) \, d\bar{r}$$

$$(33) \qquad = \frac{2^{T/2}}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} \frac{d^k}{dm^k} [m^{-T/2} I_{T/2}(m)].$$

Using the identity ([8], p. 79)

$$\frac{d}{dz} [z^{-\nu} I_\nu(z)] = z^{-\nu} I_{\nu+1}(z)$$

and changing the integration variable in (29) from p to m, we obtain finally

$$(34) \qquad \bar{E}[\bar{r}^k] = \frac{T}{2} \rho^{-T/2} \int_0^\infty m^{T/2 - 1} \exp\left(-\frac{m(1 + \rho^2)}{2\rho} \right) \frac{d^{k-1}}{dm^{k-1}} [m^{-T/2} I_{T/2+1}(m)] \, dm.$$

For $k = 1$, we have ([8], p. 386)

$$(35) \qquad \bar{E}[\bar{r}] = \frac{\rho}{1 + \frac{2}{T}}.$$

For $k = 2$, after some tedious calculation, we find

$$\bar{E}[\bar{r}^2] = \frac{1}{T + 2} + \frac{\rho^2 T(T + 1)}{(T + 2)(T + 4)}$$

$$(36)$$

$$\bar{\sigma}_{\bar{r}}^2 = \frac{1}{T + 2} \left[1 - \frac{\rho^2 T(T - 2)}{(T + 2)(T + 4)} \right]$$

We note that $\lim_{T \to \infty} \tilde{E}(\bar{r}) = \rho$ and $\lim_{T \to \infty} \tilde{\sigma}_{\bar{r}}^2 = 0$, so that at least to the extent of approximation furnished by $\tilde{R}_\rho(\bar{r})$, $\bar{r}$ is a consistent estimate of ρ.

REFERENCES

[1] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 14–33.

[2] W. J. DIXON, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[3] H. RUBIN, "On the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 211–215.

[4] R. L. ANDERSON, "Distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13.

[5] W. G. MADOW, "Note on the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308–310.

[6] B. O. PEIRCE, *A Short Table of Integrals*, Ginn and Co., 1929.

[7] G. A. CAMPBELL AND R. M. FOSTER, *Fourier Integrals for Practical Applications*, Bell Tel. Tech. Pub , 1942.

[8] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, Second Rev. ed., Cambridge University Press, 1944.

CONCERNING THE EFFECT OF INTRACLASS CORRELATION ON CERTAIN SIGNIFICANCE TESTS

By John E. Walsh

Princeton University

1. Summary. In practical applications it is frequently assumed that the values obtained by a sampling process are independently drawn from the same normal population. Then confidence intervals and significance tests which were derived under the assumption of independence are applied using these values. Often the assumption of independence between the values may be at best only approximately valid. For some cases, however, it may be permissible to assume that the correlation between each two values is the same (intraclass correlation). The purpose of this paper is to investigate the effect of this intraclass correlation on the confidence coefficients and significance levels of several well known confidence intervals and significance tests which were derived under the assumption of independence, and to extend these considerations to the case of two sets of values.

In the first part of the paper the relations given in Table I are used to compute tables which show the effect of intraclass correlation on the confidence coefficients and significance levels of the confidence intervals and significance tests listed in Table II. The second part of the paper consists of the proofs of the relations given in Table I.

2. Introduction. Let the n values $x_1, \ldots, x_n$ represent a single value of a normal multivariate population for which each of the n variables has mean μ, variance σ^2, and the correlation between each two variables is ρ. These n values will be called a correlated "sample." The values $x_1, \cdots, x_n$ and $y_1, \cdots, y_m$ are said to represent two correlated "samples" if they have a normal multivariate distribution such that the x's have mean μ, variance σ^2, correlation ρ, the y's have mean μ', variance σ'^2, correlation ρ', and the correlation between each x and y is ρ''. This paper shows that several well known quantities which have Student t, χ^2, or Snedecor F distributions when the values form random samples still have these same distributions for correlated "samples" if the quantities are multiplied by suitable constant factors, where it is to be remembered that for normal populations a correlated "sample" is a random sample if and only if $\rho = 0$ and that two correlated "samples" represent two random samples if and only if $\rho = \rho' = \rho'' = 0$. The quantities considered and the corresponding factors are listed in Table I, where $\bar{x} = \sum_1^n x_i/n$ and $\bar{y} = \sum_1^m y_a/m$. Several commonly used confidence intervals and significance tests based on these quantities and derived under the assumption of randomness are considered, and tables are computed which show how the confidence coefficients and significance levels of

these confidence intervals and significance tests vary if the values are from correlated "samples" instead of random samples. Table II contains an outline of the confidence intervals and significance tests considered. It is found that these confidence coefficients and significance levels can change noticeably when a correlated "sample" is considered. This is particularly true for the Student t-test. For example, in one case it is found that if the sample size is 32 and the significance level is .05 when $\rho = 0$, then the significance level becomes .23 for $\rho = .05$. This large change in significance level for a small change in ρ is explained by the factor given for the Student t-distribution in Table I. This shows that test results which appear to be "significant" under the assumption of randomness are not necessarily "significant" when correlation is present, even though the amount of correlation may be small. The effect of correlation on the

<div align="center">TABLE I</div>

Quantity	Distribution For Random Sample	Factor Multiplying Statistic for Correlated "Samples"
$\dfrac{(\bar{x} - \mu)\sqrt{n(n-1)}}{S} = \dfrac{(\bar{x} - \mu)\sqrt{n(n-1)}}{\sqrt{\sum\limits_1^n (x_i - \bar{x})^2}}$	Student t-distribution $g_{n-1}(t)\, dt$	$\sqrt{\dfrac{1-\rho}{1+(n-1)\rho}}$
$\dfrac{S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum\limits_1^n (x_i - \bar{x})^2$	χ^2-distribution $f_{n-1}(\chi^2)\, d\chi^2$	$\dfrac{1}{1-\rho}$
$\dfrac{\sigma'^2 S^2}{\sigma^2 S'^2} = \dfrac{\sigma'^2 \sum\limits_1^n (x_i - \bar{x})^2}{\sigma^2 \sum\limits_1^m (y_\alpha - \bar{y})^2}$	Snedecor F-distribution $h_{n-1,m-1}(F)\, dF$	$\dfrac{1-\rho'}{1-\rho}$

χ^2 and Snedecor F tests is not as great as for the Student t-test as can be seen from the factors given for the χ^2 and Snedecor F distributions in Table I.

3. Effect of intraclass correlation. The relations stated in Table I will now be used to investigate the effect of intraclass correlation on the confidence coefficients and significance levels of several common types of confidence intervals and significance tests which were derived under the assumption of random samples. The confidence intervals and significance tests considered are listed in Table II, where S^2 and S'^2 are defined in Table I. These particular confidence intervals and significance tests have the property that if α is the confidence coefficient of the confidence interval listed for a given statistic, then $1 - \alpha$ is the significance level of the significance test listed for that statistic, this relation holding whether random samples or correlated "samples" are considered. For this reason the tables given in this section will be limited to confidence coeffi-

cients; the corresponding significance levels can be obtained by using the above relation.

a.   *Student t-distribution.*   If a random sample of size n is drawn from a normal population with mean μ and variance σ^2 (denoted by $N(\mu, \sigma^2)$), a confidence interval for μ with confidence coefficient ϵ is given in Table II. If the n values form a correlated "sample", however, it follows from Table I that the corresponding confidence interval with coefficient ϵ is

$$\bar{x} - t_\epsilon S \sqrt{\frac{1 + (n-1)\rho}{n(n-1)(1-\rho)}} \leq \mu \leq \bar{x} + t_\epsilon S \sqrt{\frac{1 + (n-1)\rho}{n(n-1)(1-\rho)}}.$$

TABLE II

Statistic	Parameter Examined	Confidence Interval (Confidence Coefficient ϵ)	Significance Test (Significance Level $= 1 - \epsilon$)	Definitions of Constants
t	μ	$\bar{x} - \dfrac{t_\epsilon S}{\sqrt{n(n-1)}} \leq \mu$ $\leq \bar{x} + \dfrac{t_\epsilon S}{\sqrt{n(n-1)}}$	$\|\bar{x} - \mu\|$ $\leq t_\epsilon S/\sqrt{n(n-1)}$	$\displaystyle\int_{-t_\epsilon}^{t_\epsilon} g_{n-1}(t)\, dt = \epsilon$
χ^2	σ^2	$0 \leq \sigma^2 \leq S^2/\chi_\epsilon^2$	$\dfrac{S^2}{\sigma^2} \leq 1/\chi_\epsilon^2$	$\displaystyle\int_{\chi_\epsilon^2}^{\infty} f_{n-1}(\chi^2)\, d\chi^2 = \epsilon$
F	σ^2 σ'^2	$0 \leq \sigma^2/\sigma'^2 \leq S^2/S'^2 F_\epsilon$	$\dfrac{\sigma^2 S'^2}{\sigma'^2 S^2} \leq 1/F_\epsilon$	$\displaystyle\int_{F_\epsilon}^{\infty} h_{n-1,m-1}(F)\, dF = \epsilon$

The confidence interval given in Table II can be rewritten as

$$\bar{x} - t_\alpha S \sqrt{\frac{1 + (n-1)\rho}{n(n-1)(1-\rho)}} \leq \mu \leq \bar{x} + t_\alpha S \sqrt{\frac{1 + (n-1)\rho}{n(n-1)(1-\rho)}},$$

where

$$t_\alpha = t_\epsilon \sqrt{\frac{1 - \rho}{1 + (n-1)\rho}}.$$

Hence if $\rho < 0$, $\alpha > \epsilon$ and the confidence coefficient of the confidence interval in Table II is greater than ϵ. This means that the significance level of the corresponding significance test listed in Table II would be less than $1 - \epsilon$ so that any test result which would be significant for a random sample would also be significant for a correlated "sample" for which $\rho < 0$. If $\rho > 0$, however, $\epsilon > \alpha$ and the significance level of the test would be greater than $1 - \epsilon$. Thus a test result which would be significant for a random sample need no longer be when $\rho > 0$. The effect of positive values of ρ upon the confidence coefficient $\alpha = \alpha_t(\rho, n)$ of the confidence interval of Table II is given in Table III for the cases $\epsilon = .95$ and $.99$. Confidence intervals with unequal tails can be treated

in a similar manner. It is thus seen that the effect of correlation on the confidence coefficient increases with the sample size n, and that even a very small amount of correlation can cause a large change in α. For example, for samples of size 16 a correlation of $\rho = .05$ will change the significance level from .05 to .135; for samples of size 32 a correlation of $\rho = .05$ will change the significance level from .01 to .102, and from .05 to .23.

Confidence intervals for $\mu - \mu'$ are given by Theorem 5 of section 4. It is to be observed that if $\rho = \rho' = \rho''$ and $\sigma = \sigma'$ the confidence coefficients are independent of ρ and σ. If $m = n$, $\rho = \rho'$, $\sigma = \sigma'$, $\rho'' = 0$, however, the confidence coefficients of the confidence intervals for $\mu - \mu'$ have the values $\alpha = \alpha_t(\rho, n)$ given in Table III.

TABLE III
Values of $\alpha_t(\rho, n)$

ρ n	0	.05	.1	.2	.3	.4	.5
4	.99 .95		.983 .921	.974 .890	.961 .855	.944 .805	.920 .744
8	.99 .95		.959 .865	.913 .767	.853 .620	.790	
16	.99 .95	.865	.903 .74	.795 .64	.690 .54	.600	.515
32	.99 .95	.898 .77	.79 .68	.63			
64	.99	.79					
128	.99	.68					

b. χ^2-distribution. If a random sample of size n is drawn from $N(\mu, \sigma^2)$, a confidence interval for σ^2 with coefficient ϵ is given in Table II. If the n values form a correlated "sample", it follows from Table I that the corresponding confidence interval with coefficient ϵ is

$$0 \leqq \sigma^2 \leqq S^2/\chi_\epsilon^2(1 - \rho).$$

The confidence interval in Table II can be rewritten as

$$0 \leqq \sigma^2 \leqq S^2/\chi_\alpha^2(1 - \rho),$$

where

$$\chi_\alpha^2 = \chi_\epsilon^2/(1 - \rho).$$

Hence if $\rho < 0$, $\alpha > \epsilon$ and the significance level of the significance test given in Table II is less than $1 - \epsilon$. If $\rho > 0$, the significance level of the test is greater

than $1 - \epsilon$. The effect of positive values of ρ upon the confidence coefficient $\alpha = \alpha_{\chi^2}(\rho, n)$ of the confidence interval listed in Table II is given in Table IV for $\epsilon = .95$ and $.99$. Cases in which the lower limit of the confidence interval is not zero can be treated in a similar manner. Table IV shows that the confidence coefficient $\alpha = \alpha_{\chi^2}(\rho, n)$ decreases with the sample size n for a fixed value of ρ. Although the effect of correlation for the χ^2-distribution is not as great as for the Student t-distribution, it does cause a noticeable change in α. For example, for samples of size 16 the significance level of the test in Table II is changed from $.05$ to $.081$ if $\rho = .1$ and from $.05$ to $.13$ if $\rho = .2$. For samples of size 32 the significance level is changed from $.05$ to $.10$ for $\rho = .1$ and from $.05$ to $.19$ for $\rho = .2$.

c. Snedecor f-distribution. If two random samples, one of size n (denoted by x's) and the other of size m (denoted by y's), are drawn from $N(\mu, \sigma^2)$ and $N(\mu', \sigma'^2)$ respectively, a confidence interval for σ^2/σ'^2 with coefficient ϵ

TABLE IV

Values of $\alpha_{\chi^2}(\rho, n)$

ρ / n	0	.1	2	3	4	5.
4	.99	.988	.986	.983	.979	.971
	.95	.941	.930	.918	.900	.872
16	.99	.982	.966	.941	.890	.790
	.95	.919	.87	.79	.67	.49
32	.99	.975	.946	.867	.715	.44
	.95	.90	.81	.64	.38	.17

is given in Table II. If the values form two correlated "samples", however, it follows from Table I that the corresponding confidence interval with coefficient ϵ is

$$0 \leqq \sigma^2/\sigma'^2 \leqq \frac{S^2(1 - \rho')}{S'^2(1 - \rho)} \Big/ F_\epsilon.$$

The confidence interval in Table II can be restated as

$$0 \leqq \sigma^2/\sigma'^2 \leqq \frac{S^2(1 - \rho')}{S'^2(1 - \rho)} \Big/ F_\alpha,$$

where

$$\cdot F_\alpha = F_\epsilon(1 - \rho')/(1 - \rho).$$

Thus if $\rho = \rho'$, $\alpha = \epsilon$ and the significance level of the significance test given in Table II remains equal to $1 - \epsilon$. If $(1 - \rho')/(1 - \rho) < 1$, $\alpha > \epsilon$ and the significance level is less than $1 - \epsilon$. If $(1 - \rho')/(1 - \rho) > 1$, however, $\alpha < \epsilon$

and the significance level is greater than $1 - \epsilon$. Values of the confidence coefficient $\alpha = \alpha_F \left(\dfrac{1 - \rho'}{1 - \rho}, n, m \right)$ of the confidence interval listed in Table II are given in Table V for $\epsilon = .95$ and $.99$. Cases in which the lower limit of the confidence interval is not zero can be treated in a manner similar to that given above. Table V indicates that the effect of correlation on the confidence coefficient is not as great for $n < m$ as for $n > m$. For example, if $n = 4, m = 32$,

TABLE V

$$\textit{Values of } \alpha_F \left(\frac{1 - \rho'}{1 - \rho}, n, m \right)$$

n	m	$\dfrac{1-\rho'}{1-\rho}$ = 1	1.25	1.5	2.0
4	4	.99 .95	.987 .933	.983 .916	.975 .880
16	4	.99 .95	.978 .912	.962 .869	.917 .778
32	4	.99 .95	.975 .906	.952 .858	.896 .753
4	16	.99 .95	.987 .933	.985 .914	.977 .875
16	16	.99 .95	.973 .892	.945 .817	.858 .637
32	16	.99 .95	.919 .869	.837 .763	.628 .518
4	32	.99 .95	.987 .931	.985 .913	.977 .874
32	32	.99 .95	.960 .850	.893 .707	.675 .400

$\dfrac{1 - \rho'}{1 - \rho} = 1.25$, the significance level of the significance test given in Table II is only changed from .05 to .069, if $\dfrac{1 - \rho'}{1 - \rho} = 1.5$ from .05 to .087. If $n = 32, m = 4$, $\dfrac{1 - \rho'}{1 - \rho} = 1.25$, however, the significance level is changed from .05 to .094, if $\dfrac{1 - \rho'}{1 - \rho} = 1.5$ from .05 to .142. Also it is seen that for fixed $\dfrac{1 - \rho'}{1 - \rho}$, the effect of intraclass correlation increases with both n and m.

4. Analysis. This section contains derivations of the relations stated in the first three sections. The method used in these derivations is similar to that used in one approach to the analysis of variance and consists essentially in expressing each variable as the sum of two quantities, one of which is the same for each variable and the other of which is different for each variable.

Let $x_1, \cdots, x_n$ represent a correlated "sample", that is, have a normal multivariate distribution for which

$$E(x_i) = \mu, \qquad\qquad (i = 1, \cdots, n)$$

(1)
$$E[(x_i - \mu)^2] = \sigma^2$$

$$E[(x_i - \mu)(x_j - \mu)] = \rho\sigma^2, \qquad (i \neq j = 1, \cdots, n).$$

Write the x_i, $(i = 1, \cdots, n)$, in the form

$$x_i = \eta + \lambda\bar{\xi} + \xi_i,$$

where $\bar{\xi} = \sum_1^n \xi_i/n$ and $\eta, \xi_1, \cdots, \xi_n$ are independently distributed, η according to $N(\mu, \sigma_\eta^2)$ and the ξ_i according to $N(0, \sigma_\xi^2)$. The values of λ, σ_η^2 and σ_ξ^2 are chosen so that the $x_i = \eta + \lambda\bar{\xi} + \xi_i$ satisfy (1). It is easily proved that it is always possible to choose λ, σ_η^2 and σ_ξ^2 so that (1) are satisfied. It is to be remembered that $\rho \geq -1/(n - 1)$ for intraclass correlation. From relations (1) and $x_i = \eta + \lambda\bar{\xi} + \xi_i$ it follows that

(2)
$$E(\xi_i^2) = \sigma^2(1 - \rho), \qquad\qquad (i = 1, \cdots, n).$$

THEOREM 1. *The quantity* $\dfrac{1}{\sigma^2(1 - \rho)} \sum_1^n (x_i - \bar{x})^2$ *has a χ^2-distribution with* $n - 1$ *degrees of freedom and is distributed independently of $\bar{x}$.*

PROOF. Since the ξ_i are independently distributed according to the same normal distribution with zero mean, it follows from (2) that

$$\frac{1}{E(\xi_i^2)} \sum_1^n (\xi_i - \bar{\xi})^2 = \frac{1}{\sigma^2(1 - \rho)} \sum_1^n (x_i - \bar{x})^2$$

has a χ^2-distribution with $n - 1$ degrees of freedom and is distributed independently of $\bar{x} = \eta + (1 + \lambda)\bar{\xi}$.

THEOREM 2. $\dfrac{(\bar{x} - \mu)\sqrt{n(n - 1)}}{\sqrt{1 + (n - 1)\rho}} \bigg/ \sqrt{\sum_1^n \dfrac{(x_i - \bar{x})^2}{1 - \rho}}$ *has a Student t-distribution with $n - 1$ degrees of freedom.*

PROOF. It is easily seen from elementary considerations that $\dfrac{(\bar{x} - \mu)\sqrt{n}}{\sigma\sqrt{1 + (n - 1)\rho}}$ has the distribution $N(0, 1)$. Theorem 2 is then an immediate consequence of Theorem 1.

Up to this point a single correlated "sample" of size n has been considered. The next part of the analysis, however, will be concerned with properties which arise from the consideration of two correlated "samples."

Let $x_1, \cdots, x_n, y_1, \cdots, y_m$ have a joint normal multivariate distribution such that

(3)
$$E(x_i) = \mu, \qquad\qquad\qquad (i = 1, \cdots, n)$$
$$E(y_\alpha) = \mu', \qquad\qquad\qquad (\alpha = 1, \cdots, m)$$
$$E[(x_i - \mu)^2] = \sigma^2$$
$$E[(y_\alpha - \mu')^2] = \sigma'^2$$
$$E[(x_i - \mu)(x_j - \mu)] = \rho\sigma^2, \qquad (i \neq j = 1, \cdots, n)$$
$$E[(y_\alpha - \mu')(y_\beta - \mu')] = \rho'\sigma'^2, \qquad (\alpha \neq \beta = 1, \cdots, m)$$
$$E[(x_i - \mu)(y_\alpha - \mu')] = \rho''\sigma\sigma'.$$

Write the x_i and y_α in the form

(4)
$$x_i = \eta + \lambda_1\bar{\xi} + \lambda_2\bar{\xi}' + \xi_i$$
$$y_\alpha = \eta' + \lambda_1'\bar{\xi} + \lambda_2'\bar{\xi}' + \xi_\alpha',$$

where $\bar{\xi}' = \sum_1^m \xi_\alpha'/m$ and $\eta, \eta', \xi_1, \cdots, \xi_n, \xi_1', \cdots, \xi_m'$ are independently distributed, η according to $N(\mu, \sigma_\eta^2)$, η' according to $N(\mu', \sigma_\eta'^2)$, the ξ_i according to $N(0, \sigma_\xi^2)$, and the ξ_α' according to $N(0, \sigma_\xi'^2)$. The quantities $\lambda_1, \lambda_2, \lambda_1', \lambda_2',$ $\sigma_\eta^2, \sigma_\eta'^2, \sigma_\xi', \sigma_\xi'^2$ are chosen so that the x_i and y_α satisfy (3). It is easily verified that it is always possible to choose these quantities so that the x_i and y_α constructed in this fashion satisfy (3). In addition it follows from (3) and (4) that

(5)
$$E(\xi_i^2) = \sigma^2(1 - \rho)$$
$$E(\xi_\alpha'^2) = \sigma'^2(1 - \rho').$$

THEOREM 3. $\dfrac{1}{\sigma^2(1 - \rho)} \sum_1^n (x_i - \bar{x})^2$ *and* $\dfrac{1}{\sigma'^2(1 - \rho')} \sum_1^m (y_\alpha - \bar{y})^2$ *have χ^2-distributions with $n - 1$ and $m - 1$ degrees of freedom respectively, and are distributed independently of each other and of $\bar{x}$ and $\bar{y}$.*

PROOF. From Theorem 1 and (5) it follows that $\dfrac{1}{\sigma^2(1 - \rho)} \sum_1^n (x_i - \bar{x})^2$ and $\dfrac{1}{\sigma'^2(1 - \rho')} \sum_1^m (y_\alpha - \bar{y})^2$ have χ^2-distributions with $n - 1$ and $m - 1$ degrees of freedom respectively. That they are distributed independently of each other and of both $\bar{x}$ and $\bar{y}$ follows from (4).

THEOREM 4. $\dfrac{\sigma'^2(1 - \rho') \sum_1^n (x_i - \bar{x})^2}{\sigma^2(1 - \rho) \sum_1^m (y_\alpha - \bar{y})^2}$ *is distributed according to the Snedecor F-distribution $h_{n-1, \, m-1}(F)dF$.*

PROOF. This follows from Theorem 3.

THEOREM 5.

$$\frac{[(\bar{x} - \bar{y}) - (\mu - \mu')]\sqrt{n + m - 2}}{\sigma_1} \Bigg/ \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{\sigma^2(1 - \rho)} + \frac{\sum_1^m (y_\alpha - \bar{y})^2}{\sigma'^2(1 - \rho')}}$$

where

$$\sigma_1^2 = \frac{\sigma^2}{n} [1 + (n - 1)\rho] + \frac{\sigma'^2}{m} [1 + (m - 1)\rho'] - 2\rho''\sigma\sigma',$$

has a Student t-distribution with $n + m - 2$ degrees of freedom.

PROOF. It is easily seen from elementary considerations that $\dfrac{1}{\sigma_1} [(\bar{x} - \bar{y}) - (\mu - \mu')]$ has the distribution $N(0, 1)$. Theorem 5 then follows from Theorem 3.

The author wishes to express his appreciation to Professor John W. Tukey for valuable assistance and advice in the preparation of this paper.

ON FAMILIES OF ADMISSIBLE TESTS

By E. L. LEHMANN

University of California, Berkeley

1. Summary. For each hypothesis H of a certain class of simple hypotheses, a family F of tests is determined such that
- (a) given any test w of H there exists a test w' belonging to F which has power uniformly greater than or equal to that of w.
- (b) no member of F has power uniformly greater than or equal to that of any other member of F.

The effect on F of various assumptions about the set of alternatives are considered. As an application an optimum property of the known type A_1 tests is proved, and a result is obtained concerning the most stringent tests of the hypotheses considered.

2. Introduction. In the theory of testing simple hypotheses, if a uniformly most powerful test exists, it is the most desirable test to use. If, as is generally the case, such a test does not exist, the choice between tests none of which is "altogether better" than all the others, has to be based on information not contained in the general formulation of the testing problem. If no such additional information is available, the choice must of necessity be somewhat arbitrary.

Now although a single uniformly most powerful test exists only in exceptional cases, there will always exist a family F of tests such that
- (a) given any test w of the hypotheses H under consideration and of prescribed level of significance, there exists a test w' belonging to F which has power uniformly greater than or equal to that of w.
- (b) no member of F has power uniformly greater than or equal to that of any other member of F.

The family F is essentially unique. Arbitrariness occurs only since a test region is not uniquely determined by its power function. But since two tests with the same power function are equivalent for testing purposes, it is from the present point of view immaterial which one is included in F.

With the same restriction F is essentially the family of admissible tests, a test w being admissible if there is no test of the same level of significance which has power uniformly greater than or equal to but not identically equal to that of w. This definition differs only trivially from the one given by Wald [1, p. 15] who defines a test w to be non-admissible if there exists a w' with power everywhere greater than that of w (except at the hypothetical point).

F naturally depends on the class of alternatives considered. A restriction in the class of alternatives may (although it will not necessarily) diminish F. The family F may also be decreased by other additional information: For instance a probability distribution may be assumed for the set of alternatives, and some properties of this distribution may be presupposed.

The determination of the family F, (and a description of the power functions of the tests in F) might be considered a solution of the testing problem. The solution is not unique and hence does not provide a basis for action. This reflects the fact that additional information is needed to make possible the unique choice of a best test. On the basis of the available information, F represents the furthest reduction of the problem that seems possible. On the one hand, if the choice of test is to be made from the point of view of power, the only contestants for "best test" are the members of F. On the other hand, the available information does not give preference to any one member of F over any other unless additional principles (such as unbiasedness for instance) are introduced.

It is the purpose of the present paper to illustrate the above notions by determining F for a very simple case.

3. Determination of the family F. Let the random variable

$$E = (X_1, X_2, \cdots ; X_n)$$

have a probability density function

(1)
$$p_\theta$$

depending on parameter θ. Concerning (1) we shall make the assumptions under which Neyman [2, 3] has shown the existence of the type A_1 test of the hypothesis

(2)
$$H: \theta = \theta_0 .$$

ASSUMPTIONS:
(a) *Conditions of regularity:*
The integral

(3)
$$\int_w p_\theta(e) \, de \qquad\qquad \begin{aligned} e &= (x_1, \cdots, x_n) \\ de &= dx_1 \cdots dx_n \end{aligned}$$

extended over any region w in the sample space, admits of two successive derivatives with respect to θ under the integral sign, i.e.

(4)
$$\frac{d^k}{d\theta^k} \int_w p_\theta(e) \, de = \int_w \frac{\partial^k}{\partial\theta^k} p_\theta(e) \, de \quad \text{for} \quad k = 1, 2.$$

(b) *A differential equation:*
If

(5)
$$\varphi_\theta(e) = \frac{\partial}{\partial\theta} \log p_\theta(e)$$

$$\varphi'_\theta(e) = \frac{\partial}{\partial\theta} \varphi_\theta(e),$$

φ_{θ_0} is not identically zero, and there exist functions of θ (but independent of e), A and B, such that

(6)
$$\varphi_\theta' = A + B\varphi_\theta .$$

Under these assumptions Neyman has shown

A. that the probability density function p_θ is of the form

(7)
$$p_\theta(e) = \exp \{P(\theta) + T(e) \cdot Q(\theta) + R(e)\}$$

where Q is a monotone function with $\frac{d}{d\theta} Q(\theta) |_{\theta=\theta_0} \neq 0$ (without loss of generality we shall assume Q monotonely increasing) and

B. that the type A_1 test of the hypothesis H exists, and is given by

(8)
$$T(e) \leq c_1 , \qquad T(e) \geq c_2$$

for suitable choice of c_1 and c_2 .

In what follows we shall assume that the permissible first kind error in testing H is fixed throughout and has the value ϵ. By a test w of H we shall always mean a test of level of significance ϵ, i.e. satisfying

(9)
$$\int_w p_{\theta_0}(e) \, de = \epsilon.$$

Let us consider the family of tests

(10)
$$w(k) \colon T(e) \leq k, \, T(e) \geq f(k) ; \, k < f(k)$$

where $f(k)$ is determined by (9). It easily follows from (9) that k can take on all values from $-\infty$ to k_0 , say, where k_0 is such that

(11)
$$f(k_0) = +\infty.$$

For the family F of tests $\{w(k)\}$, $-\infty \leq k \leq k_0$ we now state

THEOREM 1. *All members of F are admissible, and if w is any admissible test not in F, there exists a member of F which has power identical with that of w.*

We first prove the

LEMMA. *Let β_w denote the powerfunction of a test w. Then if $k_1 < k_2$*

(12)
$$\beta_{w(k_1)}(\theta) < \beta_{w(k_2)}(\theta) \qquad\qquad if \; \theta < \theta_0$$
$$\beta_{w(k_1)}(\theta) > \beta_{w(k_2)}(\theta) \qquad\qquad if \; \theta > \theta_0 .$$

PROOF: Let $\bar{w}$ denote the complement of a region w. Consider the intervals

(13)
$$I = w(k_1) \cdot \overline{w(k_2)}$$
$$J = \overline{w(k_1)} \cdot w(k_2).$$

I lies entirely to the right of J. Let $\theta > \theta_0$. Then

$$(14) \qquad \frac{p_\theta(e)}{p_{\theta_0}(e)} = C(\theta) \exp \{T(e)[Q(\theta) - Q(\theta_0)]\}$$

is a strictly increasing function of T since Q is increasing. Therefore there exists a constant C such that

$$\frac{p_\theta(e)}{p_{\theta_0}(e)} \leq C \quad \text{if} \quad T(e) \text{ is in } J$$

(15)

$$C \leq \frac{p_\theta(e)}{p_{\theta_0}(e)} \quad \text{if} \quad T(e) \text{ is in } I.$$

Since

$$(16) \qquad \int_{w(k_1)} p_{\theta_0}(e) \, de = \int_{w(k_2)} p_{\theta_0}(e) \, de$$

we have

$$(17) \qquad \int_{T(e) \epsilon I} p_{\theta_0}(e) \, de = \int_{T(e) \epsilon J} p_{\theta_0}(e) \, de$$

and therefore

$$(18) \qquad \int_J p_\theta(e) \, de < C \cdot \int_J p_{\theta_0}(e) \, de = C \cdot \int_I p_{\theta_0}(e) \, de < \int_I p_\theta(e) \, de$$

from which it follows that

$$(19) \qquad \int_{w(k_2)} p_\theta(e) \, de < \int_{w(k_1)} p_\theta(e) \, de$$

which is the desired result.

PROOF OF THEOREM 1. The proof consists of several parts.

I. Let m be any real number, and assume that there exists a value of k such that

$$(20) \qquad \beta_w(\theta_0) = \epsilon$$

$$(21) \qquad \frac{d}{d\theta} \beta_w(\theta) \big|_{\theta=\theta_0} = m .$$

for $w = w(k)$. Then $w(k)$ has power uniformly greater than or equal to that of any other test satisfying (20) and (21).

For $m = 0$ this becomes Neyman's theorem stating that the type A test is also of type A_1. The proof of the theorem however is independent of the value of

$$(23) \qquad \qquad \cdot \qquad \frac{d}{d\theta} \beta_w(\theta) \big|_{\theta=\theta_0}$$

and hence carries over to arbitrary m.

II. If there exists any test satisfying (20) and (21) then there exists a number k for which $w(k)$ also satisfies (20) and (21).

To prove this let us determine, of all tests satisfying (20), the one which maximizes

$$\text{(24)} \qquad \frac{d}{d\theta} \beta_w(\theta) \mid_{\theta=\theta_0} = \int_w \frac{\partial}{\partial\theta} p_\theta(e) \mid_{\theta=\theta_0} de.$$

This can be done by means of the lemma of Neyman and Pearson [4, p. 11] which gives sufficient conditions for a region w, subject to restrictions

$$\text{(25)} \qquad \int_w f_i(e) \, de = a_i, \qquad\qquad (i = 1, \cdots, p),$$

to maximize an integral

$$\text{(26)} \qquad \int_w g(e) \, de.$$

According to this lemma the desired test is of the form

$$\text{(25)} \qquad \frac{\partial}{\partial\theta} p_\theta(e) \mid_{\theta=\theta_0} \geq a \cdot p_{\theta_0}(e)$$

provided a value of a exists for which this test satisfies (20). (25) is equivalent to

$$\text{(26)} \qquad P'(\theta_0) + T(e) \cdot Q'(\theta_0) \geq a \qquad\qquad \text{from (7)}$$

or, since $Q'(\theta_0) > 0$, to

$$\text{(27)} \qquad T(e) \geq b.$$

Thus, if a number b exists such that the test (27) satisfies (20), this test is the one maximizing (24). But such a number does exist, namely $f(-\infty)$. Therefore $w(-\infty)$ is the desired test.

Similarly it is easy to show that of all tests satisfying (20), $w(k_0)$ minimizes (24). But

$$\text{(28)} \qquad \frac{d}{d\theta} \beta_{w(k)}(\theta) \mid_{\theta=\theta_0} = \int_{T<k,\,T\geq f(k)} \frac{\partial}{\partial\theta} p_\theta(e) \mid_{\theta=\theta_0} de$$

is a continuous function of k, and therefore takes on all intermediate values, which establishes II.

III. From I. and II. we conclude that given any test w there exists a member of F which has power uniformly greater than or equal to that of w. For let w be any test of H. From the condition of regularity it follows that its power-function has a derivative at θ_0. By II. there exists a value of k such that the powerfunction of $w(k)$ has the same slope at θ_0, and from I. it follows that $w(k)$ is uniformly more powerful than w.

But from the lemma we see that none of the tests $w(k)$ is uniformly more powerful than any other. Hence all members of F are admissible, and the theorem is proved.

From the lemma and Theorem 1 we can conclude for all members of F the following optimum property:

COROLLARY 1: *Let* w *be any test, and let* w_0 *be any member of* F. *Then at least one of the two statements*

(29)
$$\beta_w(\theta) \leq \beta_{w_0}(\theta) \qquad \text{for all } \theta < \theta_0$$
$$\beta_w(\theta) \leq \beta_{w_0}(\theta) \qquad \text{for all } \theta > \theta_0$$

must hold.

The lemma and Theorem 1 also give the following result concerning most stringent tests, defined by Wald [1, p. 33].

COROLLARY 2: *There exists a uniformly most powerful of all most stringent tests. It is that unique member* w_0 *of* F *for which*

$$\operatorname*{l.u.b.}_{\theta < \theta_0} \left[\operatorname*{l.u.b.}_{w} \beta_w(\theta) - \beta_{w_0}(\theta) \right] = \operatorname*{l.u.b.}_{\theta > \theta_0} \left[\operatorname*{l.u.b.}_{w} \beta_w(\theta) - \beta_{w_0}(\theta) \right].$$

4. The effect on F of assumptions about the alternatives. Let us next consider how a restriction in the set of alternatives effects the family F. From the lemma it follows that there is no change as long as the set of alternatives contains values of θ both greater and less than θ_0. On the other hand, if the alternatives are restricted to values of θ greater than θ_0, say, the family F for testing H against these alternatives consists of only a single member, the test $w(-\infty)$, (and similarly for the other onesided case). This follows from

THEOREM 2: *Under conditions a. and b. the test* w$(-\infty)$ *is uniformly most powerful against the alternatives* $\theta > \theta_0$, *the test* w(k_0) *is uniformly most powerful against the alternatives* $\theta < \theta_0$.

PROOF: Let w be any test. By Theorem 1 there exists a number k such that

(30)
$$\beta_w(\theta) \leq \beta_{w(k)}(\theta) \qquad \text{for all } \theta.$$

From the lemma it follows that

(31)
$$\beta_{w(k)}(\theta) \leq \beta_{w(k_0)}(\theta) \qquad \text{if } \theta < \theta_0$$
$$\beta_{w(k)}(\theta) \leq \beta_{w(-\infty)}(\theta) \qquad \text{if } \theta > \theta_0.$$

Combining (30) and (31) we have the desired result.

(It is also easy to prove Theorem 2 directly from the Neyman-Pearson lemma.)

In order to illustrate how the assumption of an a priori distribution of θ together with some information about this distribution affects F, let us consider a special case of the class of hypotheses discussed so far.

Let

(32)
$$p_\theta(x_1, \cdots, x_n) = c \cdot e^{-\frac{1}{2}\Sigma(x_i - \theta)^2}$$

so that $E = (X_1, X_2, \cdots, X_n)$ is a sample from a normal distribution with unit variance and unknown mean. We want to test the hypothesis

(33)
$$H: \theta = 0.$$

We shall show that if θ has a probability density function g which is symmetric

about the origin, then the family F for testing H consists, as might be expected, of a single member, the type A_1 test.

Our problem is to find the test w satisfying

$$(34) \qquad \int_w p_0(x_1, \cdots, x_n) \, dx_1 \cdots dx_n = \epsilon$$

and which maximizes

$$(35) \qquad \int_{-\infty}^{\infty} g(\theta) \int_w p_\theta(x_1, \cdots, x_n) \, dx_1 \cdots dx_n \cdot d\theta.$$

Inverting the order of integration, which is permissible in this case, the Neyman-Pearson lemma shows the desired test to be of the form

$$(36) \qquad \int_{-\infty}^{\infty} g(\theta) p_\theta(x_1, \cdots, x_n) \, d\theta \geq a \cdot p_0(x_1, \cdots, x_n)$$

provided a value of a exists for which (36) satisfies (34). Substituting from (32), (36) becomes

$$(37) \qquad f(\bar{x}) = \int_{-\infty}^{\infty} g(\theta) e^{-\frac{1}{2}\theta^2 + n\theta\bar{x}} \, d\theta \geq a$$

where

$$(38) \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Since

$$(39) \qquad \frac{d^2}{d\bar{x}^2} f(\bar{x}) > 0$$

the region (37) is either empty, which would contradict (34), or else can be described by inequalities

$$(40) \qquad \bar{x} \leq a_1, \qquad \bar{x} \geq a_2$$

where

$$(41) \qquad f(a_1) = f(a_2)$$

the latter equation becoming, on substitution from (37)

$$(42) \qquad \int_{-\infty}^{\infty} g(\theta) e^{-(n/2)\theta^2} (e^{na_2\theta} - e^{na_1\theta}) \, d\theta = 0.$$

If g is an even function, (42) is certainly satisfied when $a_1 = -a_2$. Our test then becomes

$$(43) \qquad \bar{x} \leq -a_2, \qquad \bar{x} \geq a_2$$

which for proper choice of a_2 satisfies (34) and is the well known type A_1 test.

5. Concluding remarks. Let us consider once more a probability density function satisfying *a.* and *b.* We have seen that the family F for testing H against the alternatives $\theta \neq \theta_0$ contains an infinity of elements unless we make some additional assumptions. On the other hand, if the principle of unbiasedness is accepted, F shrinks to a single element: the type A_1 test.

But unbiasedness does not insure power. Thus conceivably some other test might be more powerful than the test chosen, everywhere except in a small one sided neighbourhood of θ_0. That this is not so is shown by Corollary 1 to Theorem 1. This remark illustrates how intuitively appealing principles and a knowledge of the family F may be used in conjunction to arrive at a choice of a satisfactory test, when not enough information is available to make the choice compelling.

Finally, it should be pointed out that although we restricted our considerations to simple hypotheses, the notions developed also apply to composite hypotheses.

<div style="text-align:center">REFERENCES</div>

[1] A. WALD, "On the principles of statistical inference". *Notre Dame Mathematical Lectures.* Number 1.

[2] J. NEYMAN, "L'estimation statistique traitée comme un problème classique de probabilité". *Conférences internationales des sciences mathématiques à Genève: Colloque d'Octobre 1937, sur le Calcul des Probabilités,* Paris, 1938.

[3] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part II". *Statistical Research Memoirs,* Vol. II, London, 1938.

[4] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part I". *Statistical Research Memoirs,* Vol. I, London, 1936.

CONDITIONAL EXPECTATION AND UNBIASED SEQUENTIAL ESTIMATION[1]

By David Blackwell

Howard University

1. Summary. It is shown that $E[f(x)\ E(y \mid x)] = E(fy)$ whenever $E(fy)$ is finite, and that $\sigma^2 E(y \mid x) \leq \sigma^2 y$, where $E(y \mid x)$ denotes the conditional expectation of y with respect to x. These results imply that whenever there is a sufficient statistic u and an unbiased estimate t, not a function of u only, for a parameter θ, the function $E(t \mid u)$, which is a function of u only, is an unbiased estimate for θ with a variance smaller than that of t. A sequential unbiased estimate for a parameter is obtained, such that when the sequential test terminates after i observations, the estimate is a function of a sufficient statistic for the parameter with respect to these observations. A special case of this estimate is that obtained by Girshick, Mosteller, and Savage [4] for the parameter of a binomial distribution.

2. Conditional expectation. Denote by x any (not necessarily numerical) chance variable and by y any numerical chance variable for which $E(y)$ is finite. There exists a function of x, the conditional expectation of y with respect to x [3, pp. 95–101, 5, pp. 41–44] which we denote, as usual, by $E(y \mid x)$ and which is uniquely defined except for events of zero probability, such that whenever $f(x)$ is the characteristic function of an event F depending only on x (i.e. $f = 1$ when F occurs and $f = 0$ when F does not occur), the equation

$$(1) \qquad E[f(x)E(y \mid x)] = E[f(x)y]$$

holds. Now if $f(x)$ is a simple function, i.e. a finite linear combination of characteristic functions, it is clear from the linearity of expectation that (1) continues to hold. Quite generally, we shall prove

THEOREM 1: *The equation* (1) *holds for every function $f(x)$ for which $E[f(x)y]$ is finite.*

To simplify notation, we write $E(z \mid x) = E_x z$ for any chance variable z. The following corollary to Theorem 1 asserts simply that the operations E_x and multiplication by $f(x)$ are commutative. This fact, which is trivially equivalent to Theorem 1, has been stated by Kolmogoroff [5, p. 50].

COROLLARY: *If $E[f(x)y]$ is finite, then $E_x[f(x)y] = f(x)E_x y$.*

PROOF OF COROLLARY: If $g(x)$ is a characteristic function, then $E(gfE_x y) = E(gfy)$ by Theorem 1. Since $E_x(fy)$ is unique, the Corollary follows.

PROOF OF THEOREM 1: Since Theorem 1 holds when $f(x)$ is a simple function and the product of a simple function and a characteristic function is a simple function, the Corollary holds when $f(x)$ is a simple function.

Now let $f(x)$ be any function for which $E(fy)$ is finite. There is a sequence of simple functions $f_n(x)$ such that $f_n(x) \to f(x)$ and $|f_n(x)| \leq |f(x)|$. For instance we may define $f_n(x) = m/n$ when $m/n \leq f(x) < (m+1)/n$, $0 \leq m \leq n^2$, $f_n(x) = m/n$ when $(m-1)/n \leq f(x) < m/n$, $0 \geq m \geq -n^2$, $f_n(x) = 0$ otherwise.

We recall the following proposition of Doob [2, p. 296]:

$$(2) \qquad\qquad | E_x y | \leq E_x | y |$$

with probability one. Then, using the Corollary (for simple functions) and (2), we have $|f_n E_x y| = |E_x(f_n y)| \leq E_x |f_n y| \leq E_x |fy|$. Also

$$(3) \qquad\qquad E(f_n E_x y) = E(f_n y).$$

Since the two sequences of functions $f_n E_x y$, $f_n y$ are bounded in absolute value by the summable functions $E_x | fy |$, $| fy |$, Lebesgue's theorem [8, p. 29] applied to (3) yields (1).

In section 3 we shall use the fact that if u is a sufficient statistic for a parameter θ and f is any unbiased estimate for θ, then $E(f \mid u)$ (which, since u is a sufficient statistic, is a function of u independent of θ) is an unbiased estimate for θ. This is obvious, since it follows from the definition of conditional expectation that the two chance variables f and $E(f \mid u)$ have the same expected value. The interesting fact is that the estimate $E(f \mid u)$ is always a better estimate for than f in the sense of having a smaller variance, unless f is already a function of u only, in which case the two estimates f and $E(f \mid u)$ clearly coincide. This is simply the fact that the variance of the regression function of f on u is not greater than the variance of f. In the case of Gaussian variables, where the regression is linear, this fact has been noted by Doob [1, p. 231].[2] Our statement is embodied in

THEOREM 2: *If $\sigma^2 y$ is finite, so is $\sigma^2 E_x y$, and $\sigma^2 E_x y \leq \sigma^2 y$, with equality holding only if $E_x y = y$ with probability one.*

PROOF: Denote by m the common expected value of y and $E_x y$. Suppose for the moment that $\sigma^2 E_x y$ is finite. By the Schwarz inequality $E[y E_x y]$ is then finite. Then $\sigma^2 y = E(y - m)^2 = E[(y - E_x y) + (E_x y - m)]^2 = E(y - E_x y)^2 + \sigma^2 E_x y$, since $E[E_x y(E_x y - m)] = E[y(E_x y - m)]$ by Theorem 1. Thus $\sigma^2 y$ exceeds $\sigma^2 E_x y$ by $E(y - E_x y)^2$, which is positive unless $y = E_x y$, i.e. y is a function of x. Thus we obtain the usual decomposition: the variance of y is the variance of the regression of y on x plus the variance of y about the regression of y on x.

To show that $\sigma^2 E_x y$ is finite, we require the following

LEMMA (SCHWARZ INEQUALITY): *If $E(f^2)$ and $E(g^2)$ are finite, then, with probability one,*

$$E_x^2(fg) \leq E_x(f^2) E_x(g^2).$$

A proof can be constructed on the usual lines by considering the function $Q(x, \lambda) = E_x(f + \lambda g)^2$. There are, however, certain measure-theoretic difficulties

---

[2] For functions of finite variance it is possible to interpret conditional expectation as a projection in Hilbert space, when the statement becomes simply the Bessel inequality.

in handling simultaneously the conditional expectations of the family of chance variables $(f + \lambda g)^2$; instead we shall give a simple direct proof based on the ordinary Schwarz inequality for integrals.

We may suppose $f \geq 0$, $g \geq 0$ with probability one, since, from (2),

$$E_x^2(fg) \leq E_x^2(|f| |g|)$$

with probability one. Unless the Lemma holds there are three positive numbers a, b, c with $a > bc$ for which the event

$$\{E_xfg > a^{\frac{1}{2}}, \qquad E_x(f^2) < b, \qquad E_x(g^2) < c\} = H$$

has positive probability. Then denoting by h the characteristic function of H and using the Schwarz inequality for integrals, we have

$$aP^2(H) \leq E^2[hE_x(fg)] = E^2(hfg) \leq E(hf^2)E(hg^2)$$
$$= E[hE_x(f^2)]E[hE_x(g^2)] \leq bcP^2(H),$$

which is impossible. This completes the proof of the Lemma.

The Lemma, with $f = y$, $g = 1$, yields $E_x^2(y) \leq E_x(y^2)$ with probability one, which implies the finiteness of $\sigma^2 E_x y$ and hence completes the proof of Theorem 2.

3. Unbiased sequential estimation. Consider a chance variable z whose distribution depends on a parameter θ. If we have an unbiased estimate $t(z)$ and a sufficient statistic $u(z)$ (not necessarily a single numerical chance variable) for θ, then, as mentioned in section 2, $v(u) = E(t | u)$ is an unbiased estimate for θ depending only on u.[3] We have shown that the variance of v is never greater than that of t, and we shall see that it is sometimes much smaller (see example II at the end of this section). The estimate obtained in this section for the parameter of a sequential process is of the v type; its importance lies in the fact that in many cases there is an unbiased estimate t (generally poor) which is a function of the first observation, and which will consequently be an unbiased estimate no matter what sequential test procedure is used.

Let $x_1, x_2, \cdots$ be a sequence of chance variables whose joint distribution is determined by an unknown point θ in a parameter space. A sequential sample (test) [9] is determined by specifying a sequence of mutually exclusive events $S_1, S_2, \cdots$, where S_i depends only on $x_1, \cdots, x_i$ and

$$(4) \qquad \sum_{i=1}^{\infty} P(S_i) = 1 \quad \text{for all } \theta.$$

The event S_i is that sampling stops after the ith observation, and (4) ensures that sampling stops eventually. Thus if we define the chance variable $n = i$ when S_i occurs, n is the size of the sample.

---

[3] It was pointed out by the referee that, strictly speaking, u does not have to be sufficient; it is necessary only that $v(u)$ be independent of θ. The author is indebted to the referee for many valuable suggestions.

Denote by u_1, u_2, $\cdots$ any sequence of chance variables such that $u_i = u_i(x_1, \cdots, x_i)$ is a sufficient statistic for estimating θ from $x_1, \cdots, x_i$. There will of course be many such sequences $\{u_i\}$, but it often happens that there is one which arises in a natural way from the sequential process; if we are sampling from a binomial population, for instance, $u_i =$ number of defectives in the first i observations is a sufficient statistic. We shall suppose that the sequential test satisfies the following condition

$$(6) \qquad\qquad S_i = W_i C(S_1 + \cdots + S_{i-1}),^4$$

where W_i is an event depending on u_i only. This condition means that when the ith observation is taken, the decision to stop at this point depends only on the ith sufficient statistic u_i. For the binomial example mentioned above, this means that the decision to stop after i observations depends only on the number of defectives observed at that stage, and not on the order in which they were observed. The Neyman criterion for u_i to be a sufficient statistic [7, 10, p. 135] shows that (6) is no restriction whatever for the sequential probability ratio test [9] since the ratio in terms of which the test is defined will be a function of u_i only.

Let t_1, t_2, $\cdots$ be any sequence of chance variables such that t_i is a function of $x_1, \cdots, x_i$; define $t = t_i$ when S_i occurs. If $E(t) = \theta$, t is said to be an unbiased estimate for θ (relative to the particular sequential test $\{S_i\}$). The theory of sequential sampling has been formulated primarily for testing hypotheses; a problem which arises naturally and often is the following: After a sequential sample has been obtained, is there an unbiased estimate for θ? Since a sample of constant size is a special case of a sequentially selected sample, we cannot hope to find unbiased estimates for arbitrary sequential samples unless such estimates exist for samples of every constant size. This is equivalent to the existence of a function $t(x_1)$ for which $E(t) = \theta$ for all θ. Our problem is to discover an unbiased estimate for θ which, when $n = i$, is a function of u_i alone. Such an estimate has been found by Girshick, Mosteller, and Savage [4] for sequential samples from a binomial population. It turns out that whenever there is any unbiased estimate at all for a particular sequential test, there is also one of the type described. Thus, if there is an unbiased estimate t for samples of fixed size N, there will be an unbiased estimate of the type described for every sequential test requiring at least N observations, since t is itself an unbiased estimate for such sequential tests.

Denote by t any unbiased estimate for θ relative to a particular sequential test $\{S_i\}$. Denote by w_i, h_i the characteristic functions of the events W_i, $C(S_1 + \cdots + S_i)$ respectively, and define $u = u_i$, $v = E(h_{i-1}t_i \mid u_i)/E(h_{i-1} \mid u_i)$ when $n = i$. To justify the definition of v we remark that the event $\{n = i, E(h_{i-1} \mid u_i) = 0\}$ has probability zero, since $qh_{i-1} \leq h_{i-1}$ with probability one, where q is the characteristic function of the event $\{E(h_{i-1} \mid u_i) > 0\}$, while

---

4 For any event A, $C(A)$ denotes the event that A does not occur.

$$E(qh_{i-1}) = E[qE(h_{i-1} \mid u_i)] = E[E(h_{i-1} \mid u_i)] = E(h_{i-1}).$$

Since u_i is a sufficient statistic for θ with respect to $x_1, \cdots, x_i$, v is a function of u and n only, independent of θ. The main result of this section is

THEOREM 3. *v is an unbiased estimate for θ.*

PROOF: We shall show that $v = E(t \mid u, n)$. This not only shows that v is an unbiased estimate for θ, but also interprets v in a very simple way and, as mentioned above, implies that the variance of v does not exceed that of t. It must be verified that for every event D depending only on n and u, $E(dv) = E(dt)$, where d is the characteristic function of D. Now $D = \sum_{i=1}^{\infty} DS_i$, and $DS_i = D_i S_i$ where D_i is an event depending only on u_i. It is sufficient, then, to show $E(d_i w_i h_{i-1} v) = E(d_i w_i h_{i-1} t)$, where d_i is the characteristic function of D_i. Now

$$E(d_i w_i h_{i-1} v) = E[d_i w_i h_{i-1} E(h_{i-1} t_i \mid u_i)/E(h_{i-1} \mid u_i)],$$

using the definition of v. The function in brackets is h_{i-1} multiplied by a function of u_i; by Theorem 1 its expectation is unaltered if h_{i-1} is replaced by $E(h_{i-1} \mid u_i)$. Thus the right member of the last equality equals

$$E[d_i w_i E(h_{i-1} t_i \mid u_i)] = E(d_i w_i h_{i-1} t_i) = E(d_i w_i h_{i-1} t).$$

We conclude with two examples:

I. BINOMIAL AND POISSON DISTRIBUTIONS. Suppose $x_1, x_2, \cdots$ are independent with identical distributions, either binomial or Poisson, with parameter θ. Then $t = x_1 \ (= t_i$ for all $i)$ is an unbiased estimate for θ, and it is well known that $u_i = x_1 + \cdots + x_i$ is a sufficient statistic for estimating θ from $x_1, \cdots, x_i$. For any sequential test satisfying (6) our unbiased estimate for θ will be

$$v = \frac{E(h_{i-1} x_1 \mid u_i = u)}{E(h_{i-1} \mid u_i = u)} = \frac{E(h_{i-1} x_1 f)}{E(h_{i-1} f)}$$

when $n = i$, $u_i = u$, where f is the characteristic function of the event $u_i = u$. Then

$$v = \frac{\sum_{j=1}^{\infty} j k_j(u, i)}{\sum_{j=0}^{\infty} k_j(u, i)} \qquad \text{for Poisson}$$

$$v = \frac{k_1(u, i)}{\sum_{j=0}^{1} k_j(u, i)} \qquad \text{for binomial}$$

where $k_j(u, i)$ denotes the number of possible sequences $x_1, \cdots, x_i$ for which $n \geq i, x_1 + \cdots + x_i = u$, and $x_1 = j$. For the binomial case, this is the estimate found in [4].

II. SAMPLES OF CONSTANT SIZE. We consider the special case where a

sample of constant size N is selected, $x_1, \cdots, x_N$ are independent with identical distributions, and the density function for x_i has the form

(7) $$p(x, \theta) = r(\theta)s(\theta)^{w(x)}q(x)$$

considered by Koopman [6][5]. Suppose further that there is an unbiased estimate $t(x_1)$ for θ. These conditions will be satisfied, for instance, if θ is the mean of a binomial, Poisson, or normal distribution, with $w(x) = t(x) = x$. Then $u_N = w(x_1) + \cdots + w(x_N)$ is a sufficient statistic. Our estimate v becomes simply $v = E[t(x_1) \mid u_N]$. Now $E[t(x_1) \mid u_N] = \cdots = E[t(x_N) \mid u_N]$, since u_N is a symmetric function of $x_1, \cdots, x_N$, which are independent with identical distributions. Consequently

$$v = E\left[\sum_{j=1}^{N} t(x_j)/N \mid u_N \right],$$

so that

$$\sigma^2(v) \leq \sigma^2\left(\sum_{j=1}^{N} t(x_j)/N\right) = \sigma^2 t(x_1)/N.$$

In the special case $w(x) = t(x) = x$, we have $v = \sum_{j=1}^{N} x_j/N$, i.e. our estimate is simply the mean of the N observations $x_1, \cdots, x_N$.

REFERENCES

[1] J. L. Doob, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.

[2] J. L. Doob, "The law of large numbers for continuous stochastic processes," *Duke Math. Jour.*, Vol. 6 (1940), pp. 290–306.

[3] J. L. Doob, "Stochastic processes with an integral-valued parameter," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 87–150.

[4] M. A. Girshick, Frederick Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems, with applications," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[5] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2 (1933).

[6] B. O. Koopman, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.

[7] J. Neyman, *Giornale dell Istituto Italiano degli Attuari*, Vol. 6 (1934), pp. 320–334.

[8] S. Saks, *Theory of the Integral*, Stechert, 1937.

[9] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[10] S. S. Wilks, *Mathematical Statistics*, Princeton Univ. Press, 1943.

[5] It has been shown by Koopman [6] that if there is a sufficient statistic satisfying certain regularity conditions, the density function for x must be of the form (7).

THE DISTRIBUTION OF THE MEAN

By E. L. Welker

University of Illinois

1. Summary. Both population and sample mean distributions can be represented or approximated by Pearson curves if the first four moments of the population are finite. Using the α_3^2, δ chart of Craig [2] to determine the Pearson curve type for the population, an analogous $\bar{\alpha}_3^2$, $\bar{\delta}$ chart is derived for the distribution of the mean. This defines a one to one transformation of α_3^2, δ into $\bar{\alpha}_3^2$, $\bar{\delta}$. The properties of this transformation are used to discuss the approach to normality of the distribution of the mean as dictated by the central limit theorem. This is facilitated by superposing on the α_3^2, δ chart the $\bar{\alpha}_3^2$, $\bar{\delta}$ charts for samples of 2, 5, and 10.

2. Introduction. For any given distribution function of a population, a method is available for finding the distribution function of the mean, when it exists, that depends on characteristic functions and the Fourier integral theorem. For example, characteristic functions have been used to show that the arithmetic means of samples from a normal population is normal, and, with minor restrictions on non-normal populations, that it is asymptotically normal. The method depends, of course, on a knowledge of the exact population distribution. Some authors have discussed the approximation of the distributions of sample means in special cases by one of the Pearson curves. It is the purpose of this paper to consider the complete range of Pearson curves as populations to be sampled, then to give the sampling distributions of the mean as approximated by the Pearson system, and to discuss the manner in which the distribution of the mean approaches the normal curve as dictated by the central limit theorem. Since the choice of a Pearson curve depends only on moment relationships, this will include the approximation of the distribution of the mean for any parent population as based on its moments. Both an algebraic and a graphic analysis will be given.

3. Semivariant and moment relationships. Denote by α_k the kth order moment of the population with zero mean and unit variance. Let λ_k be the kth order seminvariant of the population. Let $\bar{\alpha}_k$ and $\bar{\lambda}_k$ be the same parameters of the distribution of $\bar{x}$, the mean of a random sample of size N drawn from this parent population. Using properties of the seminvariants of linear functions of variables independent in the probability sense, formulas relating these parameters [1] are

$$\bar{\lambda}_k = \lambda_k N^{1-k},$$

$$\bar{\alpha}_3^2 = \bar{\lambda}_3^2 = \alpha_3^2 N^{-1},$$

and

$$\bar{\alpha}^4 = [\alpha^4 + 3(N-1)]N^{-1}.$$

4. The Pearson system of curves and the distribution of the mean. The determination of the Pearson curve will be made in accordance with the scheme discussed by C. C. Craig [2]. In this system the curve type is fixed by the moment α_3 and the constant

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}.$$
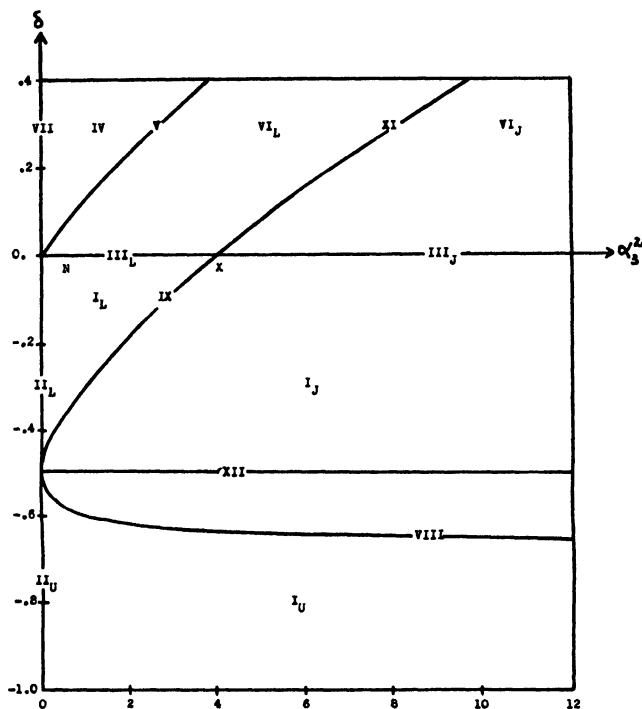
Fig. 1. The α_3^2, δ Chart for Pearson's Curves

The scheme for determining the type of curve is shown graphically in Fig. 1 in which the α_3^2, δ plane is divided into areas in which the Pearson curve types are noted. The bounding α_3^2, δ curves are

$$\delta = -1, \qquad \delta = -\tfrac{1}{2}, \qquad \delta = 0, \qquad \delta = \tfrac{2}{5}, \qquad \alpha_3^2 = 0,$$

$$\alpha_3^2 = 4\delta(\delta + 2), \quad \text{and} \quad (2 + 3\delta)\alpha_3^2 = 4(1 + 2\delta)^2(2 + \delta).$$

Let $\bar{\delta}$ denote the value of the δ function for the distribution of the mean. Then

$$\bar{\delta} = \frac{2\bar{\alpha}_4 - 3\bar{\alpha}_3^2 - 6}{\bar{\alpha}_4 + 3}.$$

In terms of moments of the parent population

$$\bar{\delta} = \frac{2\left[\dfrac{\alpha_4 + 3(N-1)}{N}\right] - 3\dfrac{\alpha_3^2}{N} - 6}{\dfrac{\alpha_4 + 3(N-1)}{N} + 3} = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3 + 6(N-1)}.$$

We see that $\bar{\delta} = \delta$ for $N = 1$, and $\bar{\delta} < \delta$ for $N > 1$. Both $\bar{\delta}$ and $\bar{\alpha}_3^2$ approach zero as N approaches infinity. These are the values of the constants for the normal function. This result is expected from the central limit theorem.

5. The $\bar{\alpha}_3^2$, $\bar{\delta}$ diagram for varying sample size. For every given population with finite moments of orders 1 through 4 there exists a Pearson curve representing or approximating its distribution. This determines a point in the α_3^2, δ plane. For a given sample size, N, there corresponds a point in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane. If the point $(\bar{\alpha}_3^2, \bar{\delta})$ is now plotted on the α_3^2, δ plane, we can determine the type of Pearson curve which is needed to approximate the distribution of x. The transformation of α_3^2, δ into $\bar{\alpha}_3^2$, $\bar{\delta}$ enables us to analyze the relationship between population distributions and distributions of $\bar{x}$. The transforms of the boundary curves in the α_3^2, δ plane will constitute an $\bar{\alpha}_3^2$, $\bar{\delta}$ chart corresponding to the one for α_3^2, δ shown in Fig. 1. In studying the approach to normality of the distribution of x, it is illuminating to superimpose this $\bar{\alpha}_3^2$, $\bar{\delta}$ chart on the α_3^2, δ chart. In order to do this, it is necessary to make certain algebraic changes in the equations.

First eliminate α_4 from the formula for δ as follows. From

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3} \quad \text{we find} \quad \alpha_4 = \frac{3\delta + 3\alpha_3^2 + 6}{2 - \delta}.$$

Substitute this in the expression for $\bar{\delta}$. Then

$$\bar{\delta} = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3 + 6(N-1)} = \frac{\delta(\alpha_3^2 + 4)}{\alpha_3^2 + 4 + 2(N-1)(2-\delta)}.$$

This formula, in conjunction with

$$\alpha_3^2 = N\bar{\alpha}_3^2$$

enables us to write the transformations of the boundary curves.

Boundary Curve	*Transformed Curve*
$\delta = -1$	$\bar{\delta} = \dfrac{-(N\bar{\alpha}_3^2 + 4)}{N\bar{\alpha}_3^2 + 4 + 6(N-1)}.$
$\delta = -\frac{1}{2}$	$\bar{\delta} = \dfrac{-(N\bar{\alpha}_3^2 + 4)}{2(N\bar{\alpha}_3^2 + 4) + 10(N-1)}.$
$\delta = 0$	$\bar{\delta} = 0.$

$$\delta = \tfrac{2}{5} \qquad\qquad \bar{\delta} = \frac{2(N\bar{\alpha}_3^2 + 4)}{5(N\bar{\alpha}_3^2 + 4) + 16(N - 1)}.$$

$$\alpha_3^2 = 4\delta(\delta + 2) \qquad \bar{\alpha}_3^2[N\bar{\alpha}_3^2 + 4 + 2\bar{\delta}(N - 1)]^2$$

$$= 4\bar{\delta}(\bar{\alpha}_3^2 + 4)[\bar{\delta}(N\bar{\alpha}_3^2 + 8N - 4) + 2N\bar{\alpha}_3^2 + 8].$$

$$(2 + 3\delta)\alpha_3^2 = 4(1 + 2\delta)(2 + \delta) \qquad [\bar{\delta}(16N + 3N\bar{\alpha}_3^3 - 4) + 2N\bar{\alpha}_3^2 + 8]$$

$$[N\bar{\alpha}_3^2 + 4 + 2\bar{\delta}(N - 1)]^2 N\bar{\alpha}_3^2$$

$$= 4[\bar{\delta}(2N\bar{\alpha}_3^2 + 10N - 2) + N\bar{\alpha}_3^2 + 4][\bar{\delta}(N\bar{\alpha}_3^2 + 8N - 4) + 2N\bar{\alpha}_3^2 + 8].$$
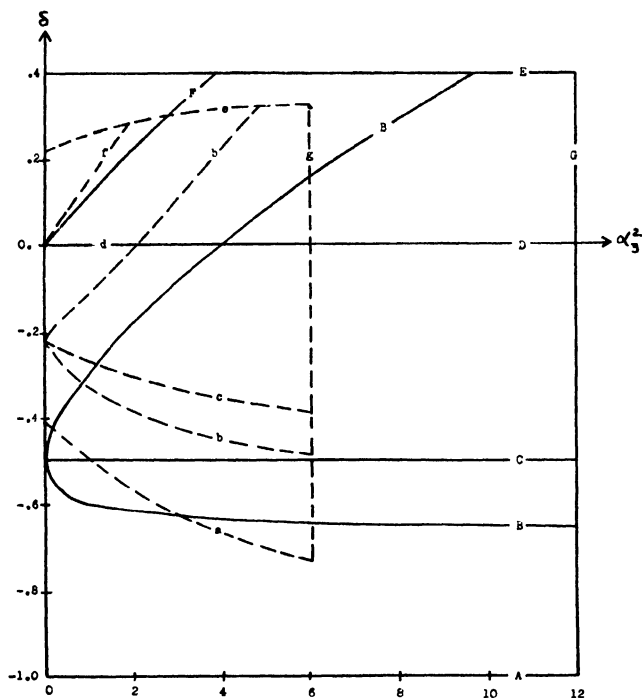
Fig. 2.   The $\bar{\alpha}_3^2$, $\bar{\delta}$ and α_3^2, δ Charts. $N = 2$

Fig. 2 shows the chart for distributions of $\bar{x}$ for $N = 2$ by dashed curves superimposed on the chart for the population shown by the solid curves, and Fig. 3 consists of the same curves for $N = 5$ and $N = 10$.   The intervals on the population values are $0 \leq \alpha_3^2 \leq 12$ and $-1 \leq \delta \leq .4$ in Fig. 2, but only part of the α_3^2 range is shown in Fig. 3.   In each case the curves for the distribution of $\bar{x}$ cover the interval for α_3^2, δ which corresponds to the entire interval shown for the population in Fig. 2.   Population curves are identified by capital letters and the corresponding curves for the distribution of $\bar{x}$ by the corresponding lower case letters.

Before discussing the Pearson curve relationships disclosed by these graphs, let us analyze some of the geometric properties of the transformation itself. Let N be considered as the parameter defining families of curves in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane corresponding to $\alpha_3^2 = $ constant and $\delta = $ constant, the systems of lines parallel to the coordinate axes. The transform of $\alpha_3^2 = k$ is $\bar{\alpha}_3^2 = k/N$, a system of lines perpendicular to $\bar{\delta} = 0$, and approaching $\bar{\alpha}_3^2 = 0$ with increasing N at the rate kN^{-2}. The line $\alpha_3^2 = 0$ is invariant under the transformation, but it is not pointwise invariant.
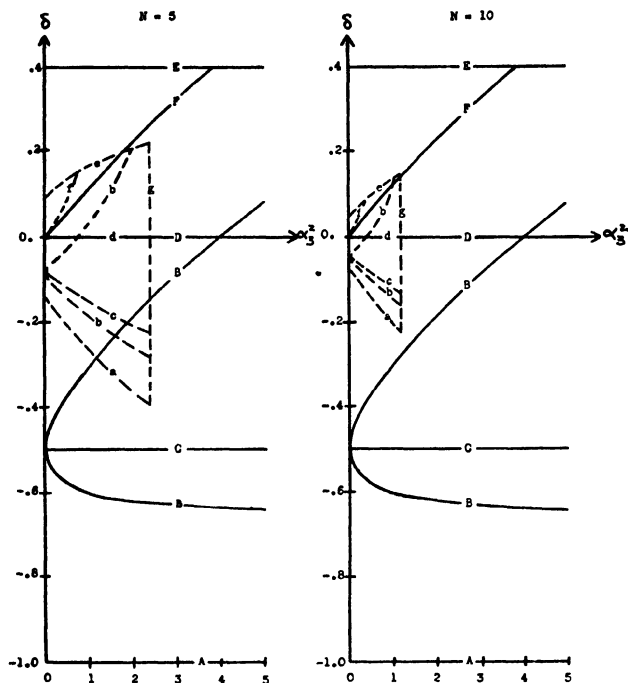
Fig. 3. The $\bar{\alpha}_3^2$, $\bar{\delta}$ and α_3^2, δ Charts

The transform of $\delta = C$ is

$$\bar{\delta} = \frac{C(N\bar{\alpha}_3^2 + 4)}{N\bar{\alpha}_3^2 + 4 + 2(N-1)(2-C)}, \text{ or } \bar{\delta} = \frac{C(\bar{\alpha}_3^2 + 4/N)}{\bar{\alpha}_3^2 + [4 + 2(N-1)(2-C)]N^{-1}}.$$

Solving for $\bar{\alpha}_3^2$, this becomes

$$\bar{\alpha}_3^2 = \frac{4C - \bar{\delta}[4 + 2(N-1)(2-C)]}{N(\bar{\delta} - C)}.$$

Except for the straight line $\bar{\delta} = 0$, obtained when $C = 0$, this is a system of rectangular hyperbolas with asymptotes

$$\bar{\alpha}_3^2 = -[4 + 2(N - 1)(2 - C)]N^{-1} \quad \text{and} \quad \bar{\delta} = C.$$

We are concerned only with the range $\bar{\alpha}_3^2 > 0$. Hence

$$-[4 + 2(N - 1)(2 - C)]N^{-1}$$

must be positive for the asymptote to show on the diagram. Since $|\delta| < 2$, and thus $|C| < 2$, the expression in brackets is necessarily positive. Hence the vertical asymptote is always outside the range of interest and will not show on the diagram. However the horizontal asymptotes, $\bar{\delta} = C$, do appear in all cases. The hyperbolas are concave downward if $C > 0$ and are concave upward if $C < 0$.

Lines of the pencil $\delta = m\alpha_3^2$ are transformed into the hyperbolas

$$\bar{\delta} = \frac{m\bar{\alpha}_3^2(N\bar{\alpha}_3^2 + 4)}{\bar{\alpha}_3^2 - 2m\bar{\alpha}_3^2(N - 1) + 4}$$

for $N > 1$. It is clear that $(0, 0)$ is the only invariant point. Every point on $\delta = m\alpha_3^2$ is transformed into a point closer to the origin, the square of the distance from the origin changing from

$$(m^2 + 1)\alpha_3^4 \quad \text{to} \; \cdot \; (m^2 + 1)\alpha_3^4 N^{-2}.$$

It is easily verified that the hyperbolas are asymptotic to

$$\bar{\delta} = \frac{mN\bar{\alpha}_3^2}{1 - 2m(N - 1)} - \frac{(N - 1)(1 + 2m)}{[1 - 2m(N - 1)]^2} \quad \text{and} \quad \bar{\alpha}_3^2 = \frac{-4}{1 - 2m(N - 1)}$$

As N approaches infinity, these asymptotes approach

$$\bar{\delta} = -\frac{\bar{\alpha}_3^2}{2} \quad \text{and} \quad \bar{\alpha}_3^2 = 0.$$

An area in quadrant one (four) in the α_3^2, δ plane is transformed into an area in quadrant one (four) in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane. The transformed area is nearer the origin.

6. Types of Pearson curves for distribution of sample means. Examination of the graphs in conjunction with the above described properties of the transformation shows the following facts regarding the distribution of means of samples drawn from populations identified by α_3^2 and δ. First consider the normal function and the three main Pearson types only.

Parent Population	Distribution of Sample Means
Normal	Normal
I_L	I_L
I_J	I_J and I_L
I_U	I_U, I_J and I_L
IV	IV
VI_L	VI_L and IV
VI_J	VI_J, VI_L and IV.

The transition types were disregarded completely in the above analysis. It is worth noting that, disregarding type X, III is transformed into III, VII into VII, II_U into II_L, never into II_U, V into IV, but never into V. Type X is transformed into type III, never into X. Others follow a similar pattern.

These moment relationships on the distribution of the mean are not sufficient conditions in general. In special cases they are, for example the normal distribution and the type III (see [3]). They do represent the best approximation curve as specified by the Pearson system. We know that in some cases, for example type II (see [3]), the distribution of means is not described by a Pearson curve. It is clear, however, that the approach to normality is indicated analytically by the transformation α_3^2, δ to $\bar{\alpha}_3^2$, $\bar{\delta}$ and is shown graphically by the $\bar{\alpha}_3^2$, $\bar{\delta}$ diagram. Skewness and kurtosis in the parent population are reflected in the distribution of the mean in small samples. A symmetric distribution of the mean requires a symmetric parent population regardless of sample size, but the degree of skewness decreases rapidly with an increasing number in the sample. The Pearson curve which approximates the distribution of $\bar{x}$ from a bell-shaped parent population is also bell-shaped. The Pearson curve approximating the distribution of $\bar{x}$ for samples of $N = 10$ (Fig. 3) is bell-shaped for any parent population with values of α_3^2 and δ within the intervals considered. For samples of 5 in the same range the approximating curve is either bell-shaped or J-shaped, but it is never U-shaped. For samples of 2, even the U-shaped distribution is possible, but only with extreme values of α_3^2 and δ. The point in the α_3^2, δ plane corresponding to the normal curve is the only invariant point in the transformation. Hence parent populations with parameters not satisfying $\alpha_3^2 = \delta = 0$ cannot yield normal distributions of sample means.

REFERENCES

[1] T. N. THIELE, "The theory of observations", *Annals of Math. Stat.*, Vol. 2 (1931), p. 206.

[2] C. C. CRAIG, "A new exposition and chart for the Pearson system of frequency curves", *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16–28.

[3] J. O. IRWIN, "On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's type II", *Biometrika*, Vol. 19 (1927), pp. 225–239.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

ON THE STUDENTIZATION OF SEVERAL VARIANCES

By B. L. Welch

University of Leeds, England

1. Introduction. In a recent paper [1] the author considered the problem of eliminating several variances simultaneously from probability statements concerning the mean of a normally distributed variable. The general situation envisaged was as follows. We supposed that we had an observed quantity y which could be assumed to be normally distributed about a population mean η with variance $\sigma_y^2 = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where the λ_i are known positive numbers and the σ_i^2 unknown population variances. It was supposed further that the data provided estimates s_i^2 of the σ_i^2 based on f_i degrees of freedom, and having the sampling distributions

$$(1) \qquad p(s_i^2) \, ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)} \exp\left\{-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right\} \left(\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right)^{\frac{1}{2}f_i - 1} d\left(\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right)$$

and that these estimates were distributed independently of each other and of y. The problem was to make statements about the magnitude of the difference $y - \eta$ which would involve explicitly only the observed variances s_i^2. The probability of the truth of the statements was also to be entirely independent of the population values σ_i^2.

The solution was given implicitly in a formal mathematical expression and a general process of developing successive terms in a series expansion was described. In the present communication a slightly different way of reaching this development is provided.

2. General method. If the f_i are large enough the ratio

$$(2) \qquad v = \frac{y - \eta}{\sqrt{\sum \lambda_i s_i^2}}$$

can be taken to be normally distributed with mean zero and standard deviation unity. This suggests that, when the f_i are not necessarily large, we might approach the matter by seeking some other function

$$(3) \qquad x = g\{s_1^2, s_2^2, \cdots, s_k^2, y - \eta\}$$

which will still be normally distributed with the same mean and standard deviation. We shall see that such a function can be found, although the method to be followed leads us first to another expression

(4)
$$y - \eta = h(s_1^2, s_2^2, \cdots, s_k^2, x)$$

which is simply the transposed form of (3). Once we have obtained h we can solve out from (4) to obtain x.

Since the distribution of y is independent of s_i^2 we have

(5)
$$p(y \mid s^2) \, dy = \frac{1}{\sqrt{2\pi \Sigma \lambda_i \sigma_i^2}} \exp \left\{ -\tfrac{1}{2} \frac{(y - \eta)^2}{\Sigma \lambda_i \sigma_i^2} \right\} dy.$$

Transforming therefore to the new variable x we have for *given* s_i^2

(6)
$$p(x \mid s^2) \, dx = \frac{1}{\sqrt{2\pi \Sigma \lambda_i \sigma_i^2}} \exp \left\{ -\tfrac{1}{2} \frac{h^2(s^2, x)}{\Sigma \lambda_i \sigma_i^2} \right\} \frac{\partial h(s^2, x)}{\partial x} \, dx$$
$$= j\{s^2, x, \Sigma \lambda_i \sigma_i^2\} \, dx \qquad \text{(say).}$$

The unrestricted distribution of x is then obtained by averaging over the joint distribution of the s_i^2. In order that x should be a unit normal deviate we must therefore have

(7)
$$p(x) = \int \cdots \int_{s^2} j\{s^2, x, \Sigma \lambda_i \sigma_i^2\} \prod_i \{p(s_i^2) \, ds_i^2\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

We have to substitute from (1) and (6) into (7) and then choose the function $h(s^2, x)$ in such a manner that the equation is satisfied whatever may be the values of the unknown σ_i^2. To evaluate the function by the methods of numerical integration is probably impracticable except perhaps in some simple special cases. A series development is, however, quite feasible.

Symbolically we can write

(8)
$$j\{s^2, x, (\Sigma \lambda_i \sigma_i^2)\} = e^{\Sigma (s_i^2 - \sigma_i^2)\partial_i} \, j\{w, x, \Sigma \lambda_i \sigma_i^2\}$$

where ∂_i denotes differentiation with respect to w_i and subsequent equation to σ_i^2. Equation (7) then integrates out to give

(9)
$$\prod_i e^{-\sigma_i^2 \partial_i} \left\{ 1 - \frac{2\sigma_i^2 \partial_i}{f_i} \right\}^{-\frac{1}{2}f_i} j\{w, x, \Sigma \lambda_i \sigma_i^2\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

i.e.

(10)
$$\Theta j\{w, x, \Sigma \lambda_i \sigma_i^2\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{(say).}$$

The operator Θ must be expanded in powers of ∂_i before it can be interpreted When this is done we find

(11)
$$\Theta = \exp \left\{ \Sigma \frac{\sigma_i^4 \partial_i^2}{f_i} + \tfrac{4}{3} \Sigma \frac{\sigma_i^6 \partial_i^3}{f_i^2} + 2\Sigma \frac{\sigma_i^8 \partial_i^4}{f_i^3} + \cdots \right\}$$

(12)
$$= 1 + \Sigma \frac{\sigma_i^4 \partial_i^2}{f_i} + \left\{ \tfrac{4}{3} \Sigma \frac{\sigma_i^6 \partial_i^3}{f_i^2} + \tfrac{1}{2} \left(\Sigma \frac{\sigma_i^4 \partial_i^2}{f_i} \right)^2 \right\} + \cdots.$$

Our procedure now is to find successive approximations to $h(s', x)$. It will be convenient to denote by $h_r(s^2, x)$ an expression which equals $h(s^2, x)$ to terms of order $1/f_i^r$. Further let $c_{r+1}(s^2, x)$ be a corrective term which when added on to $h_r(s^2, x)$ will give a result correct to terms in $1/f_i^{r+1}$. Then to this order we shall have from (6)

$$(13) \quad \sqrt{2\pi} j\{w, x, \Sigma\lambda_i\sigma_i^2\} = \frac{1}{\sqrt{\Sigma\lambda_i\sigma_i^2}} \exp\left\{-\tfrac{1}{2}\frac{h_r^2(w, x)}{\Sigma\lambda_i\sigma_i^2}\right\} \frac{\partial h_r(w, x)}{\partial x}$$
$$+ \frac{1}{\sqrt{\Sigma\lambda_i\sigma_i}} \exp\left\{-\tfrac{1}{2}\frac{x^2(\Sigma\lambda_i w_i)}{\Sigma\lambda_i\sigma_i^2}\right\} \left\{\frac{\partial c_{r+1}(w, x)}{\partial x} - \frac{x(\Sigma\lambda_i w_i)c_{r+1}(w, x)}{\Sigma\lambda_i\sigma_i^2}\right\}$$

remembering that the leading term in $h(w, x)$ is $x\sqrt{\Sigma\lambda_i w_i}$.

Hence from (10) we find

$$(14) \quad \Theta \frac{1}{\sqrt{\Sigma\lambda_i\sigma_i^2}} \exp\left\{-\tfrac{1}{2}\frac{h_r^2(w, x)}{\Sigma\lambda_i\sigma_i^2}\right\} \frac{\partial h_r(w, x)}{\partial x}$$
$$+ \frac{1}{\sqrt{\Sigma\lambda_i\sigma_i^2}} e^{-\frac{1}{2}x^2} \left\{\frac{\partial c_{r+1}(\sigma^2, x)}{\partial x} - x c_{r+1}(\sigma^2, x)\right\} = e^{-\frac{1}{2}x^2}$$

i.e.

$$(15) \quad \frac{\partial}{\partial x}\left\{e^{-\frac{1}{2}x^2}\frac{c_{r+1}(\sigma^2, x)}{\sqrt{\Sigma\lambda_i\sigma_i^2}}\right\} + \Theta \exp\left\{-\tfrac{1}{2}\frac{h_r^2(w, x)}{\Sigma\lambda_i\sigma_i^2}\right\} \frac{1}{\sqrt{\Sigma\lambda_i\sigma_i^2}} \frac{\partial h_r(w, x)}{\partial x} = e^{-\frac{1}{2}x^2}.$$

Given h_r we can therefore proceed directly to c_{r+1} and hence to h_{r+1}.

3. Application to give terms in $1/f_i$. It will be sufficient illustration of the method, if we show here how to obtain h_1 from h_0. We have from (15)

$$(16) \quad \frac{\partial}{\partial x}\left\{e^{-\frac{1}{2}x^2}\frac{c_1(\sigma^2,x)}{\sqrt{\Sigma\lambda_i\sigma_i^2}}\right\} + \left\{1 + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right\} \exp\left\{-\frac{x^2}{2}\frac{(\Sigma\lambda_i w_i)}{(\Sigma\lambda_i\sigma_i^2)}\right\} \sqrt{\frac{(\Sigma\lambda_i w_i)}{(\Sigma\lambda_i\sigma_i^2)}} = e^{-\frac{1}{2}x^2}$$

i.e.

$$(17) \quad \frac{\partial}{\partial x}\left\{e^{-\frac{1}{2}x^2}\frac{c_1(\sigma^2,x)}{\sqrt{\Sigma\lambda_i\sigma_i^2}}\right\} + \frac{(\Sigma\lambda_i^2\sigma_i^4/f_i)}{(\Sigma\lambda_i\sigma_i^2)^2} d^2 \exp\left\{-\frac{x^2 u}{2}\right\} \sqrt{u} = 0$$

where d now denotes differentiation with respect to u and subsequent equation to unity

i.e.

$$(18) \quad \frac{\partial}{\partial x}\left\{e^{-\frac{1}{2}x^2}\frac{c_1(\sigma^2, x)}{\sqrt{\Sigma\lambda_i\sigma_i^2}}\right\} = \frac{(\Sigma\lambda_i^2\sigma_i^4/f_i)}{(\Sigma\lambda_i\sigma_i^2)^2} \frac{1}{4} e^{-\frac{1}{2}x^2}(1 + 2x^2 - x^4)$$

$$(19) \quad = \frac{1}{4}\frac{(\Sigma\lambda_i^2\sigma_i^4/f_i)}{(\Sigma\lambda_i\sigma_i^2)^2} \frac{\partial}{\partial x}\left\{e^{-\frac{1}{2}x^2}(x + x^3)\right\}$$

whence

$$(20) \quad c_1(\sigma^2, x) = x\sqrt{\Sigma\lambda_i\sigma_i^2}\left[\frac{(1 + x^2)}{4}\frac{(\Sigma\lambda_i^2\sigma_i^4/f_i)}{(\Sigma\lambda_i\sigma_i^2)^2}\right].$$

Hence to the terms in $1/f_i$ we have

$$(21) \qquad y - \eta = h(s^2, x) = x\sqrt{\Sigma\lambda_i s_i^2}\left[1 + \frac{(1 + x^2)}{4}\frac{(\Sigma\lambda_i^2 s_i^4/f_i)}{(\Sigma\lambda_i s_i^2)^2}\right].$$

Solving this out for x we obtain to the same order

$$(22) \qquad x = v\left[1 - \frac{(1 + v^2)}{4}\frac{(\Sigma\lambda_i^2 s_i^4/f_i)}{(\Sigma\lambda_i s_i^2)^2}\right].$$

where v equals $(y - \eta)/\sqrt{\Sigma\lambda_i s_i^2}$. To order $1/f_i$ we may regard x as a unit normal deviate and hence determine the probability level corresponding to the observed ratio v. On the other hand if we wish to determine the value of $y - \eta$ which will lie on a given percentage level the expression (21) is the appropriate one to use.

4. Further discussion. The present development is of course basically equivalent to that given in the previous paper. Indeed if we integrate (10) or (15) out with respect to x we arrive immediately at the formulae which were then obtained and which were illustrated by calculating terms to order $1/f_i^2$. In fact when calculating higher order terms it seems best to do this integration before carrying out the operation Θ. The object of the present note is really to stress the fact that we are simply finding a function of the observations and of $y - \eta$ which is distributed as a unit normal deviate, whatever the values the true σ_i^2 may chance to possess.

Finally, the remarks following equation (7) above should be somewhat amplified. The equation asserts that the distribution of any arbitrary function x, defined by (3), is

$$(23) \qquad p(x) = \int \cdots \int \frac{1}{\sqrt{2\pi\Sigma\lambda_i\sigma_i^2}}\left\{\exp - \frac{1}{2}\frac{h^2(s^2, x)}{\Sigma\lambda_i\sigma_i^2}\right\}\frac{\partial h(s^2, x)}{\partial x}\prod_i\{p(s_i^2)\,ds_i^2\},$$

where $h(s, x)$ is the function obtained by solving out (3) for $y - \eta$. On carrying out the integrations in (23) we shall in general obtain $p(x)$ as a function of x and σ_i^2. Our argument is that if h be chosen properly the σ_i^2 will disappear from $p(x)$, and x will appear only in the form of the unit normal probability function.

To find $h(s, x)$ by a direct process of numerical integration would appear to involve in the first instance the choice of a net-work of points for x and s_i^2. Suppose the range of x is covered by n_x points and the range of s_i^2 by n_i points. We may then as an approximation look on our task as that of finding the $(n_x\pi n_i)$ values of $h(s^2, x)$ corresponding to this network. Since (23) is to be true for all x and σ_i^2, we can take in turn n_i values of σ_i^2, and then (23) can be replaced by $(n_x\pi n_i)$ simultaneous equations (it would be necessary to use some formula expressing $\partial h(s^2, x)/\partial x$ in terms of values of $h(s^2, x)$ at discrete values of x or conceivably this may be avoided if we work with the integrated form). With a proper choice of the points for x, s_i^2, and σ_i^2, we might expect to evaluate the series $h(s^2, x)$ to any required degree of accuracy, but clearly as a general process to be used over a whole range of values f_i this approach would be too laborious.

It may indeed be queried whether theoretically, with an indefinitely fine network of points, we shall be led to a unique function $h(s^2, x)$ with the common sense properties, which, from general statistical considerations, we know it should have in order to be acceptable. As with integral equations of a simpler character, the passage from a discrete network to a continuum may raise problems, but it is the author's opinion that the infinite ranges of x and s_i^2 give us the freedom which we require in the solution.

The author, however, prefers to approach the problem from the numerical behavior of the series, of which (15) gives the general terms. Here the practical issue appears to be to investigate the relation between the magnitude of the last term retained and the f_i. The author hopes in a further paper to give some results of an investigation of this character and also some tables facilitating the calculation of $h(s^2, x)$.

REFERENCE

[1] B. L. WELCH, "The generalization of 'Student's' problem when several different population variances are involved". *Biometrika*, Vol. 34 (1947), pp. 28–35.

PROBABILITY SCHEMES WITH CONTAGION IN SPACE AND TIME[1]

BY FÉLIX CERNUSCHI[2] AND LOUIS CASTAGNETTO

Harvard University

1. Summary. In many natural assemblies of elements, the probability of an event for a given element depends not only on the intrinsic nature of that particular element, but also on the states of some or all of the rest of the elements belonging to the same assembly. On the basis of this general idea of "contagion" some urn schemes are developed in this paper in which one has contagious influence in space and time. The most interesting result found is that in general the points of convergence of the probability of the assembly are given by some of the roots of an equation $p = f(p)$ and that some of these roots, between zero and one, represent stable states of the assembly, or points of convergence, and others represent unstable ones, or points of divergence. The two neighboring roots, (if they are single), of a root representing a point of convergence are unstable values of the probability. Consequently, under certain conditions, the limiting probability may be made to have a finite jump by changing the initial probability by an arbitrarily small amount. The concrete cases developed in this paper can be considerably extended by similar methods by assuming more complicated and general assemblies and laws of contagion.

---

[1] On the suggestion of the referee, some parts of the original paper were deleted and some mathematical simplifications were introduced.

[2] Research Associate at Harvard Astronomical Observatory and Guggenheim Fellow.

2. Introduction. In the known probability schemes of contagion of Eggenberger and Polya [1], Greenwood and Yule [2], Lüders [3], Neyman [4], Feller [5] and others [6], as well as in Markoff chains different ways are considered in which the previous results in a definite series of trials may influence the probabilities of the future ones. All of these schemes consider possible influences of the results of the different trials along the time axis; and consequently might be called schemes of contagion in one dimension and one direction.

In many natural assemblies of individuals or elements, the probability of an event per individual or element depends not only on the intrinsic nature of the considered element but also on the states of the rest of the elements belonging to the same assembly.

The purpose of this paper is to develop some simple schemes with urns in which there is a contagious influence in space and time and to show some of their consequences. The method which we have used to treat certain concrete cases could be applied to more complicated assemblies and laws of influence in space and time.

3. Scheme of a closed assembly of urns in two dimensions. Let us consider a set of N urns arranged on a closed surface in such a way that each one of them is surrounded by m others. Let each urn contain a finite number of black and white balls. In this paper the probability associated with an urn will refer to the probability of obtaining a white ball if a single ball is drawn at random from the urn. We shall assume that the initial probabilities are equal for all of the urns and that the following law of influence holds: When, after a collective trial, one finds that the ball drawn from a certain arbitrary urn, taken as the central one, is white and that the corresponding results of the m surrounding urns give l white and s black balls, one multiplies the probability of obtaining a white ball out of the central urn by the factor $\alpha_{1,1}^l \alpha_{1,2}^s$; if the ball drawn from the central urn were black, without changing the given results of the surrounding urns, one multiplies the considered probability by the factor $\alpha_{2,2}^l \alpha_{2,1}^s$. Under the specified conditions, it is easily seen that the probability of obtaining a white ball from a definite urn at the $i + 1$ trial will be, by considering all the possible alternatives:

$$
(1) \quad p_{i+1} = m! \sum_{j=0}^{m} \frac{1}{j!(m-j)!} [p_i^2(p_i\alpha_{1,1})^{m-j}(q_i\alpha_{1,2})^j + p_i q_i(p_i\alpha_{2,1})^{m-j}(q_i\alpha_{2,2})^j]
$$

$$
= f(p_i) = p_i^2(p_i\alpha_{1,1} + q_i\alpha_{1,2})^m + p_i q_i(p_i\alpha_{2,1} + q_i\alpha_{2,2})^m,
$$

where:

$$
p_i + q_i = 1.
$$

Consequently p_i either converges to a root of the equation $p = f(p)$ or tends to infinity. As a probability greater than one or smaller than zero has no meaning,

we have to study the function $y = f(p)$ between zero and one. In (1) we have given an implicit form for $y = f(p)$, corresponding to a particular case of influence; by changing the law of influence we change the function $f(p)$. In general one can find graphically the roots of equation $p = f(p)$ by plotting $y = f(p)$ and $y = p$ and by determining the intersections of these two lines in the range $0 \leq p \leq 1$. Later we shall give the values of these roots for some concrete examples. From what we have shown it follows that if, for the considered assembly of urns and for especially chosen values of the parameters of interconnection and initial probabilities, the probability tends to some equilibrium value, this must be a root of the equation $p = f(p)$. As we shall see later, the roots in the range $0 \leq p \leq 1$ may represent stable or unstable states of the assembly.

Let us consider now a general method for finding the explicit form of the function $f(p)$ corresponding to laws of influence similar to the one used by Polya.

Assume that the trial i results in the drawing of l white balls and s black balls from the m urns surrounding the central one. Then we add lw_1 white and sb_1 black balls to the central urn if the result of the central urn was white, and lw_2 white and sb_2 black balls if it was black. It is easy to show that under these conditions the probability in the trial $i + 1$ is related to the probability in trial i by the following formula:

$$
\begin{aligned}
(2) \quad p_{i+1} = p_i \int_0^1 t^{N_i - W_i - 1} &\left[t_1 \frac{\partial}{\partial t_1} t_1^{W_i} (p_i t_1^{w_1} + q_i t_2^{b_1})^m \right]_{t_1 = t_2 = t} dt \\
+ (1 - p_i) \int_0^1 t^{N_i - W_i - 1} &\left[t_1 \frac{\partial}{\partial t_1} t_1^{W_i} (p_i t_1^{w_2} + q_i t_2^{b_2})^m \right]_{t_1 = t_2 = t} dt
\end{aligned}
$$

where W_i and N_i are the number of white balls and the total number of balls, respectively, in the central urn before trial i. Relation (2) permits us to study several interesting schemes. It is easy to see that all the possible schemes which can be represented by relations of type (2) give only values of the probability in the interval zero and one; and consequently we do not need to make the restriction in the analysis of the equation $p = f(p)$ that was necessary in the previous scheme, represented by equation (1).

For the case $w_1 = b_1 = c_1$, $w_2 = b_2 = c_2$, we obtain from (2)

$$
(3) \qquad p_{i+1} = p_i \frac{W_i + mp_i c_1}{N_i + mc_1} + (1 - p_i) \frac{W_i + mc_2 p_i}{N_i + mc_2}.
$$

If $c_1 = c_2$, (3) gives

$$
(4) \qquad\qquad\qquad p_{i+1} = p_i.
$$

If one takes $c_1 = k_1 N_i$ and $c_2 = k_2 N_i$ (3) becomes

$$
(5) \qquad p_{i+1} = p_i \frac{p_i + mk_1 p_i}{1 + mk_1} + (1 - p_i) \frac{p_i + mk_2 p_i}{1 + mk_2} = f(p_i)
$$

and the equation $p = f(p)$ has, in this case, the roots 0 and 1.

When $w_1 = b_2 = k_1 N_i$ and $b_1 = w_2 = k_2 N_i$, one has to replace $l_1(\partial/\partial l_1)$ by $l_2(\partial/\partial l_2)$ in the second term of (2); then if we take $m = 2$,

$$(6) \quad p_{i+1} = \frac{p_i + 2k_1}{1 + 2k_1} + p_i q_i \left(\frac{p_i}{1 + 2k_2} + 2 \frac{p_i + k_1}{1 + k_1 + k_2} - 3 \frac{p_i + 2k_1}{1 + 2k_1} \right) = f(p_i).$$

In particular, if $k_1 = k_2 = k$, one obtains

$$(7) \quad p_{i+1} = \frac{1}{1 + 2k_1} [4kp_i^2 - (4k - 1)p_i + 2k] = f(p_i),$$

and the solutions of the equation $p = f(p)$ are $p = \frac{1}{2}$ and 1. By considering the behavior of $y = f(p)$ one finds that the stable solution is given by the root $\frac{1}{2}$; consequently if one starts with any value of $0 < p < 1$ the probability tends to the limiting value $\frac{1}{2}$. If $k_1 = 0$, $k_2 \neq 0$, by simple calculations, one obtains from (6) that the solutions of $p = f(p)$, in this case, are zero and one.

The equation $p = f(p)$, as given by (6), always has the solution 1. In order to have the other two roots real, one has to satisfy:

$$(8) \quad \begin{aligned} k_1(1 + 2 k_2) (2 + k_1 + 3 k_2)^2 &\geq 4(1 + k_1 + k_2) \\ & [(k_1 + k_2)^2 + 2(k_1 - k_2) - 4 k_2^2]. \end{aligned}$$

A simple and interesting application of relation (2) is for the case of two urns, characterized by $m = 1$. From (2) we obtain:

$$(9) \quad p_{i+1} = p_i^2 \left(\frac{p_i + k_1}{1 + k_1} + \frac{1 - p_i}{1 + k_2} \right) + (1 - p_i)p_i \left(\frac{p_i + k_3}{1 + k_3} + \frac{1 - p_i}{1 + k_4} \right) = f(p_i)$$

where

$$w_1 = k_1 N_i, \quad b_1 = k_2 N_i, \quad w_2 = k_3 N_i, \quad b_2 = k_4 N_i.$$

The equation $p = f(p)$, as given by (9) has the roots 0 and 1; and one may fix the value of the third root by conveniently choosing the values of the parameters.

Applying (2) for an arbitrary value of m and integrating by parts, it is seen that in general the equation $p = f(p)$ is of degree $m + 2$ and consequently, by choosing appropriate values for the parameters k_1, k_2, k_3, k_4, each of which may be between -1 and ∞, one can expect several roots in the range $0 \leq p \leq 1$. One can easily generalize our relation (2) for cases in which w_1, w_2, b_1, b_2 are given functions of the probability p_i. Even in this most general case it is simple to see that one would have a recursion formula of the type $p_{i+1} = f(p_i)$ and, as in the elementary cases which we have considered, the points of equilibrium of the closed assembly of urns will be given by those solutions, in the range $0 \leq p \leq 1$, of the equation $p = f(p)$, where the derivative of $y = f(p)$ is negative. Consequently the two neighboring roots, if they are single, of a root representing a point of convergence are unstable values of the probability. Therefore, under certain conditions, the limiting probability may take a finite jump if the initial probability is changed by an arbitrarily small amount. This is, we think, the most important consequence of the contagion schemes that we propose. We

consider that many actual cases of contagion could be better understood by schemes of the type that we are studying.

Let us consider now some simple cases of relation (1). If we take

$$\alpha_{1,1} = \alpha_{2,2} = \alpha_1 \quad \alpha_{1,2} = \alpha_{2,1} = \alpha_2 \quad \text{and} \quad m = 2,$$

representing a closed ring of urns, one obtains:

$$
\begin{aligned}
(10) \qquad p_{i+1} &= p_i^2(\alpha_1 p_i + \alpha_2 q_i)^2 + p_i q_i(\alpha_2 p_i + \alpha_1 q_i)^2 \\
&= p_i^2 + (p_i^2 - p_i^3)\left[(\alpha_1 + \alpha_2)^2 - 4\,\alpha_1^2\right] = f(p_i).
\end{aligned}
$$

The equation $p = f(p)$, corresponding to this recursion formula, always has the solution $p = 0$. The other two solutions are given by

$$(11) \qquad P_{1,2} = \tfrac{1}{2}\left[1 \pm \sqrt{\frac{4 - (\alpha_1 + \alpha_2)^2}{4\alpha_1 - (\alpha_1 + \alpha_2)^2}}\,\right].$$

These roots will be between 0 and 1 when

$$(12) \qquad \begin{array}{c} 2 < \alpha_1 + \alpha_2 \\[4pt] 1 > \alpha_1 < \alpha_2 \end{array} \quad \text{or (12')} \quad \begin{array}{c} 2 > \alpha_1 + \alpha_2 \\[4pt] 1 < \alpha_1 > \alpha_2 \end{array}.$$

We would have $P_1 > 0$ and $P_2 < 0$ if

$$(13) \qquad \begin{array}{c} 2 < \alpha_1 + \alpha_2 \\[4pt] 1 < \alpha_1 < \alpha_2 \end{array} \quad \text{or (13')} \quad \begin{array}{c} 2 > \alpha_1 + \alpha_2 \\[4pt] 1 > \alpha_1 > \alpha_2, \end{array}$$

and $P_1 = P_2$ when

$$(14) \qquad \alpha_1 + \alpha_2 = 2, \qquad \alpha_1 \neq 1.$$

Let us now study the general behavior of (10). For the conditions (12') we have:

$$(15) \qquad p_{i+1} - p_i = a^2 p_i(p_i - P_1)(p_i - P_2)$$

where

$$a^2 = 4\,\alpha_1^2 - (\alpha_1 + \alpha_2)^2 > 0.$$

If $0 < P_1 < P_2$, one obtains from (15) by use of elementary algebra:

$$(16) \qquad \left|\frac{p_{i+1} - p_i}{P_1 - p_i}\right| = a^2 p_i \,|\,(P_2 - p_i)\,| \le \frac{a^2 P_2^2}{4} \le 1.$$

Consequently if $p_1 > P_2$ the sequence p_i increases monotonically. Otherwise p_{i+1} will lie between P_1 and p_i and will tend to P_1 without ever reaching the other side of this point. In a similar way it is possible to prove the convergence to a constant for the most general equations of the type $p = f(p)$ when they have roots between zero and one.

Let us give some numerical results. For $\alpha_1 = 0.95$ and $\alpha_2 = 1.1$, from (10) one obtains: $P_1 = 0.1$ and $P_2 = 0.9$. It is easily seen that, in this case, if

$0 < p_1 < 0.1$, the limiting value of p_i will be zero; if $p_1 > 0.1$, the limiting value will be 0.9. The interesting point is that if the initial probability is in the neighborhood of 0.1, an infinitesimal change in its value may produce a finite change in the stable limiting probabilities; and that for the initial probability equal 0.1 one would have an unstable equilibrium of the system. This consideration shows why it is important to know how the probability p_i converges towards a certain point. As we have previously shown, the points of convergence are roots of the eq. $p = f(p)$ but there roots which are not points of convergence.

Similar reasoning could be applied to more complicated systems belonging to our general scheme of contagion. Consequently, the most important result is not that the considered assembly may have a probability tending to some value in the range $0 \le p \le 1$, but that under certain conditions the limiting probability may jump from one value to another by changing the initial probability by an arbitrarily small amount.

REFERENCES

[1] F. EGGENBERGER AND G. POYLÁ, *Zeits. fur Ang. Math. und Mech.* Vol. 3 (1923), p. 279.
[2] M. GREENWOOD AND G. U. YULE, *Roy. Stat. Soc. Jour.*, Vol. 83 (1920), p. 255.
[3] R. LUDERS, *Biometrika*, Vol. 26 (1934), p. 108.
[4] J. NEYMAN, *Annals of Math. Stat.*, Vol. 10 (1939), p. 35.
[5] W. FELLER, *Annals of Math. Stat.*, Vol. 14 (1943) p. 389.
[6] F. CERNUSCHI AND E. SALEME, *Anales Soc. Cientifica Argentina*, Vol. 138 (1944), p. 201.

FITTING CURVES WITH ZERO OR INFINITE END POINTS

BY EDMUND PINNEY

Oregon State College

The problem of determining a suitable equation to fit an empirically determined curve over a given interval has been of great importance in statistical work, in experimental science, and in engineering technology. Since infinitely many types of equations may be made to fit the data with required accuracy, the choice of a "suitable" type of equation depends on the qualitative nature of the empirical curve, on the use to which the equation is to be put, and upon considerations of simplicity.

As a function type, the polynomial has, because of its simplicity, been enormously useful. The function type studied here is a little more general than the polynomial type, being particularly useful in the case of empirical curves that become zero or infinity at one or both ends of the interval.

Without loss of generality the interval in which the equation is to fit the curve may be taken as $0 \le x \le 1$. It is assumed that, by numerical means or otherwise, a finite set of moment $\mu_m = \int_0^1 yx^m \, dx$ may be computed, y being the ordinate of the empirical curve.

The problem to be considered here is that of determining a function $f(x)$ of the form

$$(1) \qquad f(x) = x^{\alpha}(1 - x)^{\beta} \sum_0^n a_p x^p, \qquad R(\alpha) > -1, \qquad R(\beta) > -1$$

such that

$$(2) \qquad \int_0^1 f(x)x^m \, dx = \mu_m$$

as m ranges from zero to the number of the highest moment known. $f(x)$ is then an approximation to y which may be written

$$(3) \qquad y \approx f(x).$$

THEOREM $1°$. *Given a finite set of moments $\mu_0, \mu_1, \mu_2, \cdots, \mu_n$, and given that $R(\alpha) > -1, R(\beta) > -1$, define*

$$(4) \quad S_p(\alpha, \beta) = \frac{\Gamma(p + \alpha + 1)}{\Gamma(p + \alpha + \beta + 1)} \sum_0^p {}_m \binom{p}{m} \frac{\Gamma(m + p + \alpha + \beta + 1)}{\Gamma(m+\alpha+1)} (-)^m \mu_m,$$

$$
a_k^{(n)} = \frac{(-)^k}{k! \, \Gamma(k + \alpha + 1)}
$$
$$(5)$$
$$
\cdot \sum_p^n \frac{(2p + \alpha + \beta + 1)\Gamma(p + k + \alpha + \beta + 1)}{(p - k)! \, \Gamma(p + \beta + 1)} S_p(\alpha, \beta),
$$

$$(6) \qquad f(x) = x^{\alpha}(1 - x)^{\beta} \sum_0^n {}_k \, a_k^{(n)} x^k.$$

Then $f(x)$ will satisfy (2) for $m = 0, 1, \cdots, n$.
    $2°$. *If, in addition to $1°$, μ_{n+1} is known and α and β satisfy*

$$(7) \qquad S_{n+1}(\alpha, \beta) = 0,$$

then $f(x)$ will satisfy (2) for $m = n + 1$ also.
    $3°$. *If, in addition to $1°$ and $2°$, μ_{n+2} is also known, and if α, β also satisfy*

$$(8) \qquad S_{n+2}(\alpha, \beta) = 0,$$

then $f(x)$ will satisfy (2) for $m = n + 2$ as well.
    PROOF. Let $P_m^{(\alpha, \beta)}(z)$ be the Jacobi polynomial of order m defined in terms of the hypergeometric function by

$$(9) \qquad P_m^{(\alpha, \beta)}(z) = \binom{m + \alpha}{m} F(-m, m + \alpha + \beta + 1; \alpha + 1; \tfrac{1}{2} - \tfrac{1}{2}z).$$

Let $P_m^{(\alpha, \beta)}(1 - 2\mu)$ symbolically represent the expression gotten by substituting μ_k for x_k in the expansion of the polynomial $P_m^{(\alpha, \beta)}(1 - 2x)$. There exist numbers $A_{m,q}$ such that

$$(10) \qquad x^m = \sum_q^m A_{m,q} P_q^{(\alpha, \beta)}(1 - 2x).$$

Also

(11) $$\mu_m = \sum_0^m A_{m,q} P_q^{(\alpha,\beta)}(1 - 2\mu).$$

For $R(\alpha) > -1, R(\beta) > -1$, define

(12) $$f(x) = x^\alpha(1 - x)^\beta \sum_0^n \frac{(2p + \alpha + \beta + 1)p!\,\Gamma(p + \alpha + \beta + 1)}{\Gamma(p + \alpha + 1)\Gamma(p + \beta + 1)}$$
$$\times P_p^{(\alpha,\beta)}(1 - 2\mu)P_p^{(\alpha,\beta)}(1 - 2x).$$

Then by (10), for $m = 0, 1, \cdots, n$,

$$\int_0^1 f(x)x^m \, dx = \sum_0^m {}_p \frac{(2p + \alpha + \beta + 1)p!\,\Gamma(p + \alpha + \beta + 1)}{\Gamma(p + \alpha + 1)\Gamma(p + \beta + 1)} P_p^{(\alpha,\beta)}(1 - 2\mu)$$
$$\times \sum_0^m {}_q A_{m,q} \int_0^1 x^\alpha(1 - x)^\beta P_q^{(\alpha,\beta)}(1 - 2x)P_p^{(\alpha,\beta)}(1 - 2x) \, dx.$$

By the orthogonality of the Jacobi polynomials, [1; §4.3],

$$\int_0^1 f(x)x^m \, dx = \sum_0^n {}_p A_{m,p} P_p^{(\alpha,\beta)}(1 - 2\mu).$$

By (11),

$$\int_0^1 f(x)x^m \, dx = \mu_m, \qquad (m = 0, 1, \cdots, n).$$

It follows from (2) that $f(x)$ as defined in (12) is the $f(x)$ of (1). It remains to be shown that (12) may be expressed in the form (4)–(6).

From (9),

(13) $$P_p^{(\alpha,\beta)}(1 - 2x) = \frac{\Gamma(p + \alpha + 1)}{\Gamma(p + \alpha + \beta + 1)}$$
$$\times \sum_0^p {}_m \frac{(-)^m}{m!(p - m)!} \frac{\Gamma(p + m + \alpha + \beta + 1)}{\Gamma(m + \alpha + 1)} x^m,$$

so by (4),

(14) $$P_p^{(\alpha,\beta)}(1 - 2\mu) = \frac{1}{p!} S_p(\alpha, \beta).$$

Inserting (13) and (14) into (12),

$$f(x) = x^\alpha(1 - x)^\beta \sum_0^n {}_p \frac{2p + \alpha + \beta + 1}{\Gamma(p + \beta + 1)}$$
$$\times \sum_0^p {}_k \frac{(-)^k}{k!(p - k)!} \frac{\Gamma(p + k + \alpha + \beta + 1)}{\Gamma(k + \alpha + 1)} x^k S_p(\alpha, \beta)$$
$$= x^\alpha(1 - x)^\beta \sum_0^n {}_k \frac{(-)^k x^k}{k!\,\Gamma(k + \alpha + 1}$$
$$\times \sum_k^n {}_p \frac{(2p + \alpha + \beta + 1)\Gamma(p + k + \alpha + \beta + 1)}{(p - k)!\,\Gamma(p + \beta + 1)} S_p(\alpha, \beta),$$

$$= x^{\alpha}(1 - x)^{\beta} \sum_{0}^{n} {}_k a_k^{(n)} x^k,$$

by (5), so the $f(x)$ of (12) may be expressed in the form (4)–(6), and part $1°$ of the theorem is established.

If (7) holds, by (5), $a_k^{(n+1)} = a_k^{(n)}$ for $k = 0, 1, \cdots, n$, and $a_{n+1}^{(n+1)} = 0$. Therefore, in (6),

$$f(x) = x^{\alpha}(1 - x)^{\beta} \sum_{0}^{n+1} {}_k a_k^{(n+1)} x^k,$$

and by part $1°$, for the case in which n is replaced by $n + 1$, it follows that (2) holds for $m = n + 1$, so part $2°$ is established. The establishment of part $3°$ is essentially the same.

In applying this theorem to the problem of empirical curve fitting, it follows from (6) that the constants α and β should differ from zero only if the empirical curve approaches zero or infinity at one or both of its endpoints. With this in mind the following rules may be stated:

CASE A. If, in the empirical curve, $f(0) \neq 0$ or ∞, and $f(1) \neq 0$ or ∞, set $\alpha = \beta = 0$, and let n be one less than the number of moments that it is desired to fit.

CASE B. If $f(0) = 0$ or ∞ and $f(1) \neq 0$ or ∞, set $\beta = 0$ and determine α from (7), n being two less than the number of moments that it is desired to fit.

CASE C. If $f(0) \neq 0$ or ∞ and $f(1) = 0$ or ∞, set $\alpha = 0$ and determine β from (7), n being two less than the number of moments that it is desired to fit.

CASE D. If $f(0) = 0$ or ∞ and $f(1) = 0$ or ∞, determine both α and β from the two equations (7) and (8), n being three less than the number of moments that it is desired to fit.

It may happen that these processes cannot be carried out, or at least cannot be conveniently carried out. If this is the case, α or β may be set arbitrarily and n taken as one unit higher than before, or both α and β may be set, and n taken as two units higher than before.

In Case D, above, the solution of equations (7) and (8) may often prove difficult, making it advisable to follow the suggestions of the last paragraph. In certain special cases, however, their solution is not difficult.

Suppose, for example, the moments satisfied the equations

$$(15) \qquad \mu_m = \sum_{0}^{m} {}_q \binom{m}{q} (-)^q \mu_q, \qquad m = 0, 1, \cdots.$$

If this is substituted into (4), and the order of summation reversed, on making use of the identity

$$(16) \qquad \sum_{0}^{n} {}_p \binom{n}{p} \frac{\Gamma(p + \alpha)}{\Gamma(p + \nu)} (-)^p = (-)^n \frac{\Gamma(\alpha)\Gamma(\alpha - \nu + 1)}{\Gamma(\alpha - \nu - n + 1)\Gamma(n + \nu)},$$

one obtains

$$(17) \qquad S_p(\alpha, \beta) = (-)^p S_p(\beta, \alpha).$$

Therefore

(18) $$S_{2p+1}(\alpha, \alpha) = 0.$$

When n is an integer, either $n + 1$ or $n + 2$ is odd. Therefore when (15) holds, one of either (7) or (8) will be satisfied identically if we take $\beta = \alpha$. The other may then be solved for α.

As an example, suppose one had the moments $\mu_0 = 1$, $\mu_1 = \frac{1}{2}$, $\mu_2 = \frac{7}{24}$, $\mu_3 = \frac{3}{16}$, $\mu_4 = \frac{31}{240}$, and wished to obtain an $f(x)$ such that $f(0) = 0$, $f(1) = 0$. In this case $n = 2$, and (15) is satisfied. It follows that (7) is satisfied identically when $\beta = \alpha$, and (8) gives

$$\frac{\Gamma(2\alpha + 5)}{\Gamma(\alpha + 1)} + 4\,\frac{\Gamma(2\alpha + 6)}{\Gamma(\alpha + 2)}\left(-\frac{1}{2}\right) + 6\,\frac{\Gamma(2\alpha + 7)}{\Gamma(\alpha + 3)}\left(\frac{7}{24}\right)$$
$$+ 4\,\frac{\Gamma(2\alpha + 8)}{\Gamma(\alpha + 4)}\left(-\frac{3}{16}\right) + \frac{\Gamma(2\alpha + 9)}{\Gamma(\alpha + 5)}\left(\frac{31}{240}\right) = 0.$$

This easily reduces to

$$1 - 4\,\frac{\alpha + 5/2}{\alpha + 1} + 7\,\frac{(\alpha + 5/2)(\alpha + 3)}{(\alpha + 1)(\alpha + 2)}$$
$$- 6\,\frac{(\alpha + 5/2)(\alpha + 7/2)}{(\alpha + 1)(\alpha + 2)} + \frac{31}{240}\,\frac{(\alpha + 5/2)(\alpha + 7/2)}{(\alpha + 1)(\alpha + 2)} = 0,$$

which reduces to the quadratic

$$4\alpha^2 - 6\alpha + 5 = 0,$$

from which

(19) $$\alpha = \beta = 3/4 \pm (1/4)\sqrt{11}\,i.$$

These may be substituted into (4)–(6) to complete the solution.

REFERENCE

[1] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Colloquium Pub., No. 23, 1939.

————◆————

CONSISTENCY OF SEQUENTIAL BINOMIAL ESTIMATES

By J. Wolfowitz

Columbia University

The notion of consistency of an estimate, introduced by R. A. Fisher, applies to a sequence of estimates which converge stochastically, with boundlessly increasing sample size, to the parameter (or parameters) being estimated. Each estimate is a function of a sample of observations, the number in each sample being determined independently of the observations themselves. In sequential estimation, on the other hand, the number of observations is itself a chance

variable, determined by the sequence of observations and the application to them of a rule which may be part of a sequential test. In what follows we shall consider that the operation of sequential estimation *is* associated with a sequential test.[1]

The advantage of using consistent estimates is such as to suggest extension of the idea of consistency to sequential estimation. In the present paper we shall be concerned only with the estimation of a binomial probability (p, say). The obvious extension is that a sequence of estimates, each with its associated test, is consistent if the estimates converge stochastically to p.

Since the number of observations required by a sequential test is a chance variable, a parallel to the classical sequence of samples of increasing size would be a sequence of sequential tests whose average (in some sense) sample sizes increase without limit. It seems reasonable to associate only such a sequence of estimates with this sequence of tests as will converge stochastically to p, i.e., be consistent.

Let z be a chance variable which takes the distinct values c_1 and c_2 with probabilities p, $0 < p < 1$, and $q = 1 - p$, respectively. Let $z_1, \cdots, z_n$ be a sequence of independent observations on z which terminates with the nth according to the specific sequential test under consideration. Denote by x and y, respectively, the number of observations c_2 and c_1 in this sequence. Then x, y and $n = x + y$ are all chance variables. The couple $g = (x, y)$ is called a boundary point of index n (see [1]). The sequence of observations which terminates at g is called a path. Let $k(g)$ denote the number of paths which terminate at g, and let $k^*(g)$ denote the number of these paths whose first observation is c_1. The "points" on the various paths together with all the points g constitute the "region" under discussion.

Let $P\{n = j\}$ denote the probability of the relation in braces. If

$$\sum_{j=1}^{\infty} P\{n = j\} = 1,$$

the region is called closed. Only closed regions will be considered below, so that this assumption will henceforth be made without explicit formulation. It has been shown by Girshick, Mosteller, and Savage [1], that $p(g) = k^*(g)/k(g)$ is an unbiased estimate of p for any closed region R, i.e.,

$$\Sigma \, p(g)k(g)p^y q^x \equiv p,$$

where the summation takes place over all the boundary points g of R. For many important regions this estimate is the unique unbiased estimate.

Let there be given an infinite sequence of sequential tests with each of which we associate the estimate $p(g)$. Consider the ith one of these, and let n_{0i} be the smallest number of observations required for a decision, i.e., n_{0i} is the smallest

---

[1] Really all that is required is a rule for terminating the observations such that its region R is closed (see below). However, we defer to conventional statistical usage in referring to "tests."

value of j for which $P\{n = j\} \neq 0$. The theorem proved below asserts that if n_{0i} approaches infinity with i the estimate $p(g)$ converges stochastically to p. To put it in other words: if T_1, T_2, $\cdots$ is the sequence of tests, and ϵ_1 and ϵ_2 are arbitrarily small positive numbers, there exists a positive number $J(\epsilon_1, \epsilon_2)$ such that, for all T_i such that $i > J$,

$$P\{| p(g) - p | > \epsilon_1\} < \epsilon_2 ,$$

when $n_{0i} \to \infty$. An important example of such a sequence is that of the Wald sequential binomial tests [2] obtained as follows: Let $\alpha_1, \alpha_2, \cdots, \alpha_i \cdots$ and $\beta_1, \beta_2, \cdots, \beta_i \cdots$, be two sequences of positive numbers all of which are less than $\frac{1}{2}$ and which approach zero as $i \to \infty$. Let p_0 and p_1, $0 < p_0 < p_1 < 1$, be two fixed numbers,

$$c_1 = \log \frac{p_1}{p_0}, \qquad c_2 = \log \frac{(1 - p_1)}{(1 - p_0)}, \qquad Z_j = \sum_{k=1}^{j} z_k .$$

Finally let the rule for terminating the process of drawing observations be as follows for the ith test T_i: The process of drawing observations terminates at the smallest integer n for which either

$$Z_n \geq \log \frac{1 - \beta_i}{\alpha_i} \quad \text{or} \quad Z_n \leq \log \frac{\beta_i}{1 - \alpha_i} .$$

Since $(1 - \beta_i)/\alpha_i \to \infty$ and $\beta_i/(1 - \alpha_i) \to 0$ while c_1 and c_2 are constant, it is evident that the hypothesis of the theorem is satisfied.

The property of being unbiased is not generally considered an indispensable characteristic of an optimum estimate, while consistency is generally so regarded. Our theorem shows that $p(g)$ enjoys the latter property with respect to important sequences of sequential tests.

THEOREM: *Let $T_1, \cdots, T_i, \cdots$ be a sequence of sequential binomial tests. For the ith test T_i let n_{0i} be the smallest integer such that $P\{n = n_{0i}\} \neq 0$. Finally let $n_{0i} \to \infty$ as $i \to \infty$. Then $p(g)$ converges stochastically to p as $i \to \infty$.*

PROOF: For typographic simplicity we shall use n_0 as the designation of the generic element of the sequence $n_{01}, n_{02}, \cdots$. No confusion will be caused thereby.

Let $n' = n_0 - 1$, and $\delta_1 > 0$ and $\delta_2 > 0$ be arbitrarily small fixed numbers. Let $k'(g)$ be the number of paths which end at the point g and are such that $| y'/n' - p | < \delta_1$, where y' is the number of observations c_1 among the first n' observations. We then have

LEMMA 1. *For n_0 sufficiently large*

(1)
$$\sum_{g \epsilon B} k'(g) p^y q^x > 1 - \delta_2$$

where B is the set of boundary points of R.

PROOF: Consider the totality $\{h\}$ of all points $h = (x', y')$, with $x' + y' = n'$. Here x' and y' denote, respectively, the number of observations c_2 and c_1 in the sequence of the first n' observations on z. Let $k_0(h)$ denote the number of paths

to h. Let C denote the set of points h such that $|\ y'/n' - p\ | < \delta_1$. If n_0 is large enough we have, by the law of large numbers,

$$\sum_{h \in C} k_0(h) p^{y'} q^{x'} > 1 - \delta_2 .$$

Let $k(h, g)$ be the number of paths from h to g. From Theorem $2'$ of [3] it follows that

(A) $$\sum_{g \in B} k(h, g) p^y q^x = p^{y'} q^{x'}.$$

Also from the definitions of the various symbols involved it readily follows that

$$k'(g) = \sum_{h \in C} k_0(h) k(h, g).$$

Hence

$$\sum_{g \in B} k'(g) p^y q^x = \sum_{g \in B} \left(\sum_{h \in C} k_0(h) k(h, g) \right) p^y q^x = \sum_{g \in B} \left(\sum_{h \in C} k_0(h) k(h, g) p^y q^x \right)$$

$$= \sum_{h \in C} k_0(h) \left(\sum_{g \in B} k(h, g) p^y q^x \right) = \sum_{h \in C} k_0(h) p^{y'} q^{x'} > 1 - \delta_2 .$$

This proves Lemma 1.

Let $\xi(g) = [k(g) - k'(g)]k(g)$. Thus $\xi(g)$ is a chance variable, being a function of the chance point g.

LEMMA 2. Let δ_3 and δ_4 be arbitrarily small positive numbers. For n_0 sufficiently large

(2) $$P\{\xi(g) \leqq \delta_3\} > 1 - \delta_4 .$$

PROOF: If (2) were not true, we would have

(3) $$E \frac{k'(g)}{k(g)} = \Sigma\ k'(g) p^y q^x \leq (1 - \delta_4) + (1 - \delta_3)\delta_4 = 1 - \delta_3 \delta_4 .$$

Choose the δ_2 of Lemma 1 so that $\delta_2 < \delta_3 \delta_4$. For some large value of n_0 we would then have a contradiction between (1) and (3). This proves the lemma.

Let g be any boundary point. Consider any path whose y' is such that $|\ y'/n' - p\ | < \delta_1$; let us call such a path one of type T. Consider the terminal sequence S of this path,

$$S : z_{n_0}, z_{n_0+1}, \cdots, z_n$$

This sequence, together with $g = (x, y)$, uniquely determines y'. Any permutation of y' elements c_1 and $n' - y' = x'$ elements c_2 may serve as the initial sequence of n' observations of a path which terminates at g and has the terminal sequence S. For no boundary point is of index smaller than n_0, so that under permutation of the first n' observations a path remains a path, i.e., the process of taking observations will not terminate prematurely as a result of the permuting of the elements. Of these permutations a proportion y'/n' begin with the element c_1. We deal in this manner with all the different terminal sequences of the paths of

type T which end at g. Let $k^{*\prime}(g)$ be the number of these which begin with c_1. We obtain

LEMMA 3. *For all g such that $k'(g) \neq 0$*

$$\left| \frac{k^{*\prime}(g)}{k'(g)} - p \right| < \delta_1 .$$

Putting Lemmas 2 and 3 together we have

LEMMA 4. *As $n_0 \to \infty$, $k^{*\prime}(g)/k(g)$ converges stochastically to p.*

Now it follows in a manner similar to that of Lemma 2 that, as $n_0 \to \infty$, $k^{*\prime}(g)/k^*(g)$ converges stochastically to one. This, together with Lemma 4, proves the theorem.

REFERENCES

[1] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems, with applications," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[2] A. WALD, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[3] J. WOLFOWITZ, "On sequential binomial estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 489–492.

BOOK REVIEWS

Mathematical Methods of Statistics. *Harald Cramér.* Uppsala, Sweden: Almqvist and Wiksell, 1945. pp. xvi, 575. (Princeton, N. J.: Princeton University Press, 1946. $6.00)

REVIEWED BY WILL FELLER

Cornell University

This book represents a contribution of a novel kind to the statistical literature and will render valuable services both as textbook and reference book. Of its three parts the first one (134 pages) is entitled *Mathematical Introduction* and develops the necessary formal mathematical tools. The second part (186 pages) is devoted to *Random Variables and Probability Distributions*, that is to say, to a chapter of the modern theory of probability. The third, and main, part of the book (some 233 pages) is entitled *Statistical Inference*. Ordinarily these three topics would require consultation of three or more books, and these would rarely be found on the same shelf. However, the masterly exposition succeeds in creating the impression of natural unity and harmony. The ideas are developed with elegance and apparent ease as if the line of presentation followed a well explored path. The uninitiated will not notice how unconventional the treatment is and how the very selection of topics depends on the author's scientific personality.

It is hardly necessary to point out that Cramér's book fills an urgent need. The emergence of statistical theory and methodology as an exact science, firmly grounded in mathematical probability, is only of recent date. Its rapid development went hand in hand with an extraordinary increase of the number and importance of its various applications. Under such circumstances there was naturally little time for an exposition of the theoretical foundations and ramifications. Modern statistical inference has its roots in the classical limit theorems of probability. Now classical probability used to consist of a bewildering collection of special and mutually uncorrelated problems; unified guiding principles and methods are a rather new development and have not yet found expression in the textbook literature. The original investigations are usually written in an exceedingly abstract language and the existing close ties to applications are not apparent. Consequently, there is no easy access either to probability or statistics and it is often difficult to establish whether, or to what extent, various assertions have actually been proved. The present book therefore closes a serious gap in the literature and will greatly facilitate both teaching and research.

Of the 12 chapters of the *Mathematical Introduction* 9 are devoted to the theory of measure and integration. The antiquated theory of the so-called Riemann integral (kept alive by elementary textbooks) considered only point functions $y = f(x)$, where the independent variable is a point. The temperature at a given point or the velocity at a given moment are typical examples. Many mathematical considerations simplify greatly if from the very beginning also set

functions $y = F(A)$ are introduced, where the independent variable is a set. Typical examples are mass in mechanics, the amount of heat or of electricity, area or wealth of a geographic region, and the probability of events (i.e. sets in sample space). The Lebesgue-Stieltjes theory frees the concept of integral from artificial devices and reduces it to the natural notion of mean values with respect to set functions. In a simile, believed to be due to Lebesgue, the Riemann integral corresponds to the procedure of a grocer who computes the day's receipts by actually adding the several amounts in the order as they had come in. The Lebesgue procedure imitates the more intelligent grocer who orders his cash in piles of notes and coins of equal denomination and counts them. The analogy with the customary procedure of computing mathematical expectation is clear. The Lebesgue-Stieltjes integral is conceptually simpler than the Riemann integral and can be presented in as simple a way with rigor adequate for elementary textbooks. It has become an indispensable tool in probability, statistics, physics, and other applied fields. Since it has, unfortunately, not found its way into calculus textbooks, physicists are compelled to use the less flexible notion of the Dirac δ-function, and the formal mathematical apparatus in general becomes unnecessarily clumsy. It is a curious anomaly that so many calculus textbooks profess to be written with a view to applications and yet completely disregard the most obvious practical needs and that the teaching of practical mathematics should remain uninfluenced by the great developments of the last fifty years.

In such circumstances the chapters on integration will be particularly welcome to statisticians as probably the only place in the literature where they will find easy access to the theory. Of course, this exposition leads far beyond what the average statistician will require under ordinary circumstances and beyond the necessary prerequisites of the main body of the book. Of the 88 pages roughly half can be omitted at first reading in accordance with detailed instructions given in the Preface. The remaining half will form a valuable reference book for theorems and tools used occasionally in connection with more delicate parts of statistical theory. The mathematical introduction contains also a chapter on Fourier integrals (characteristic functions), one on matrices and quadratic forms, and finally miscellaneous complements such as orthogonal polynomials, Euler's summation formula, beta and gamma functions, etc.

The title to the second part, *Random Variables and Probability Distributions*, is the same as that of the author's well-known Cambridge Tract of 1937. Both start with a discussion of the foundations along axiomatic lines. The new treatment does not differ essentially from the old one, but some changes are introduced which are regrettable in the reviewer's opinion (in particular axiom 3). Otherwise there is practically no overlap between the two expositions. The 1937 booklet devoted much space to the asymptotic expansions connected with the central limit theorem which are due to the author himself. This topic is not touched upon in the present book. This is a judicious procedure since the 1937-booklet is generally accessible (although at present sold out). Instead we now find a detailed study of some univariate distributions such as χ^2, Student's t,

Fisher's z, the Pearson system, etc., none of which were mentioned in the Cambridge tract. Similarly, there is now a section on correlation and regression, and the normal distributions in several variables. The theory of probability is developed only to the extent of the formal theory of distribution functions. This implies that even so important a notion as stochastic convergence is treated only summarily while the strong law of large numbers falls completely outside the framework of the book. This is regrettable inasmuch as the strong law is of greater importance than the classical weak law (whose fame rests essentially on a classical misunderstanding). It should be mentioned that this second part of the book contains some 39 well chosen illustrative exercises the solution of which is left to the reader.

In the main part of the book, entitled *Statistical Inference*, the outer form changes inasmuch as the text there is accompanied by numerous practical examples. However, the exposition remains mathematical in nature and the main emphasis rests on exact formulations; much attention is paid to the establishment of the precise conditions of validity of the individual theorems, their logical interrelations and their connections with general probability. The expert will find many minor and major improvements in formulations and proofs. They are too numerous to be listed here. Suffice it to point out, as a typical example, the theorem on pp. 426–27 concerning the limiting form of the χ^2 distribution with estimated parameters; this theorem appears to be more general than usually stated and also the proof seems to be novel. The topics treated in the statistical part of the book will be seen from the following list of titles to the chapters. 25. Preliminary Notions on Sampling. 26. Statistical Inference (general orientation). 27. Characteristics of Sampling Distributions (moments, semi-invariants, corrections for grouping, etc.). 28. Asymptotic Properties of Sampling Distributions (moments, extreme values, range, etc.). 29. Exact Sampling Distributions (degrees of freedom, Student, Fisher, correlation and regression coefficients, partial and multiple correlations, generalized variance, etc.). 30. Tests of Goodness of Fit and Allied Tests (treating mostly applications of χ^2). 31. Tests of Significance for Parameters. 32. Classification of Estimates (sufficient, efficient and asymptotically efficient estimates; minimum variance, etc.). 33. Methods of Estimation (method of moments, maximum likelihood, χ^2-minimum methods). 34. Confidence Regions. 35. General Theory of Testing Statistical Hypotheses. 36. Analysis of Variance. 37. Some Regression Problems. There follow tables of the normal distribution, the χ^2 and the t-distributions, and a long list of references.

If an expression of wishes for a second edition were permitted, most statisticians would probably give first choice to non-parametric and sequential tests. It is needless to point out that the latter became public only after completion of the Swedish edition of the present book

Even this short account will show the extremely wide range of topics and theories covered in the book, from abstract integration to randomized experiments. They are all presented with uniform lucidity. The exposition through-

out is formal, and yet inspiring, rigorous and yet never pedantic. It will serve as an example worthy of imitation and is an achievement on which the author deserves our sincere congratulations.

———————

The Advanced Theory of Statistics. Vols. I and II. *Maurice G. Kendall.* London: C. Griffin and Co., Ltd. Vol. I. Second ed. revised, 1945; pp. xii, 457, 50 shillings. Vol. II. 1946; pp. viii, 521; 42 shillings.

REVIEWED BY M. S. BARTLETT

Cambridge University and *The University of North Carolina*

With the recent appearance of the second volume, it is now possible to review as one work this comprehensive treatise. To quote the author's opening remarks to the Preface to Volume I: "The need for a thorough exposition of the theory of statistics has been repeatedly emphasized in recent years. The object of this book is to develop a systematic treatment of that theory as it exists at the present time." An outline of the contents, which in the two volumes make up just on a thousand pages, will indicate that this formidable task has been squarely faced by the author, who, when a tentative co-operative venture of writing such a treatise was upset by the outbreak of the war, continued alone with the project.

Volume I contains sixteen chapters. The first six introduce the concept of frequency distributions via observational data on groups and aggregates, and their mathematical representation (Ch. 1), measures of location and dispersion (Ch. 2) and moments and cumulants in general (Ch. 3), characteristic functions (Ch. 4), and ending with a description of the standard distribution functions, such as the binomial, Poisson, hypergeometric and normal distributions, and the Pearson and Gram-Charlier systems. The next section opens with probability (Ch. 7) and proceeds to sampling theory (Chs. 8–11), including a chapter (Ch. 10) on exact sampling distributions, many of the standard sampling distributions being used in this chapter to illustrate the mathematical methods available for obtaining sampling distributions. Chapter 11 deals with the general sampling theory of cumulants, including a useful reference list of formulae and a demonstration, due to the author, of the validity of Fisher's combinatorial rules for obtaining these formulas. The section concludes with a chapter on the Chi-square distribution and some of its applications. The last four chapters of Volume I deal with association and contingency, correlation, including partial and multiple correlation, and rank correlation; this last chapter being a comprehensive treatment including comparatively recent results of the author.

It will be convenient to list also the contents of Volume II before any critical comment on either volume. The first section of the second volume comprises four chapters on the theory of estimation, including a derivation of the properties of the maximum likelihood estimate (Ch. 17) and separate chapters on Fisher's theory of fiducial probability and Neyman's theory of confidence intervals. The

second main section, according to the author's remarks in the preface to Volume II, deals with the theory of statistical tests and comprises chapters 21, 23, 24, 26, 27 and 28; of these after an introductory chapter (Ch. 21) on tests of significance, chapters 23 and 24 cover analysis of variance, chapters 26 and 27 give a fairly detailed account of the general theory of significance-tests originated by Neyman and Pearson and Chapter 28 deals with the recently developed techniques of multivariate analysis. The remaining chapters are 22 on regression, 25 on the design of sampling enquiries, and Chapters 29 and 30 on time-series, another subject in which the author has himself taken an active interest. Finally, there are two appendices, A consisting of a few addenda to Volume I, and B an extensive bibliography of theoretical statistical papers.

The volumes are attractively printed; and each chapter concludes with a useful collection of examples for the reader.

In any comprehensive treatment of a wide subject there can be no clearly defined order of presentation; nevertheless, the author's order of chapters in Volume II and in particular his inclusion of analysis of variance among the chapters on the theory of statistical tests is a little puzzling, and the reviewer's preference would have been to see this important subject treated earlier, together with regression analysis, and their link with the classical method of least squares more firmly outlined. Incidentally, there appears to be no mention of the Fourier analysis of observational data except in its relation to periodogram analysis (Ch. 30). This change of order would perhaps also have allowed a shift forward of Chapter 25 on the design of sampling enquiries, and a more compact section on multiple correlation, culminating with the chapter on multivariate analysis before the chapters on the general theory of statistical inference were begun.

Another arrangement of rather doubtful value in Volume II is the allocation of separate chapters to fiducial probability and to the theory of confidence intervals. The problem of how to deal with a field which is still a battleground is admittedly not an easy one, and this particular one is an embarrassment at present to many teachers, but it may be questioned whether strict impartiality is the best answer. To take a hypothetical example, there would seem to be no particular virtue in a textbook which expounded, in parallel, statistical methods of inference using direct probabilities and the method of "inverse probability", leaving the reader to decide at the end which he should adopt.

The most criticizable arrangement, however, occurs in Volume I with the late and rather scanty treatment of probability in Chapter 7. To begin with examples of statistical data is sound, but since the whole conceptual model erected to deal with such data is based on probability theory, it does not seem sufficient for a reader who "feels keenly on the subject" to do as the author suggests in the Preface and read Chapters 7 and 8 after Chapter 1. Even if he does so, he will find no very clear exposition of the statistical theory of probability,—no mention, for example, of the laws of large numbers, whether for simple dichotomies or for entire continuous distribution functions, that show how the conceptual model adequately corresponds with the empirical notions of "in the long run" or "for

a large enough sample". The actual arrangement, moreover, leads to an apparently rather arbitrary treatment of theorems on limiting distributions; the First Limit Theorem, which deals with the equivalence of the limits of distribution function and corresponding characteristic function sequences, is given in the chapter on characteristic functions (Ch. 4), and the Central Limit Theorem, dealing with the convergence to normality of a sum of n independent random variables, is given in the chapter on probability.

In the proof of the second part of the First Limit Theorem, dealing with the conditions under which a sequence $\phi_n(t)$ of characteristic functions determine the limiting distribution function $F(x)$, the author has not yet corrected an error that occurred in Cramér's original version, which Kendall follows (section 4.12). Correct conditions for convergence of the distribution function sequence $F_n(x)$ to $F(x)$ (at all continuity points of F) are convergence of the characteristic function sequence to $\phi(t)$ for all real t, uniformly in at least some finite t interval (cf. H. Scheffé, *Math. Reviews*, Vol. 6 (1945), p. 89).

Another proof in Volume I which appears to need clarification is the geometrical derivation of the distribution of the multiple correlation coefficient in the case of a non-zero true correlation (section 15.21). The blunt statement is made, following equation (15.51), that the sample correlation coefficient R and an angle ψ (defined in the text) are independent, a statement which is incorrect. However, if the logic of Fisher's original derivation is examined, it turns out that the relation of R and ψ is only required when the true correlation is zero; under such conditions R and ψ are independent.

In Volume II there is a sentence requiring correction and amplification in the derivation (in the case of zero true canonical correlations) of the sampling canonical correlation distribution (section 28.30). The sentence "Consider the distribution for a given value of t_{ij} and $z_{ij} \cdots$" should be corrected to read "Consider the distribution for a given value of $t_{ij} + z_{ij} \cdots$". Some justification that the distribution is independent of $t_{ij} + z_{ij}$ is then still needed.

There is inevitably, owing to the time the book was written, no mention of sequential analysis, the sampling technique developed during the war by Wald and others and only recently "derestricted". Again, in chapter 18, where the work of Aitken and Silverstone on unbiased estimates with minimum variance is referred to, the simple inequality connecting the variance of any unbiased estimate with Fisher's information function throws an interesting new light on this aspect of the estimation problem (see, for example, H. Cramér, *Mathematical Methods of Statistics*, section 32.3, or C. R. Rao, *Bulletin Calcutta Math. Soc.*, Vol. 37 (1945), p. 81), but was not known to the author when this chapter was written. Such omissions are merely an indication of the developing nature of the subject, and it is hoped they can be remedied in later editions. There is, however, especially in Volume II, an occasional impression of patchiness in the treatment not altogether excusable on such grounds. This can perhaps be illustrated from the last chapter, a valuable contribution to the still-growing subject of time-series, but where the importance of some known results does not always

seem sufficiently stressed; in particular, the Wiener-Khintchine relation between the periodogram and correlogram is noted (section 30.68) as "an interesting relation", whereas it is a fundamental relation in the modern method of approach to time-series, giving much deeper insight into the correct interpretation of classical periodogram analysis.

These criticisms, which could be extended to cover minor errors and misprints, are not intended to detract seriously from what is a remarkable achievement. An excellent sense of proportion has been maintained throughout between mathematical theory and illustrative discussion and examples. This makes this treatise, if both the breadth and level of the subject matter are taken into account, at present unique. It will be an indispensable reference book to every teacher and advanced student of the theory of statistics.

Sequential Analysis of Statistical Data: Applications. Prepared by the Statistical Research Group, Columbia University for the Applied Mathematics Panel, National Defense Research Committee, Office of Scientific Research and Development. SRG Report 255, Revised; AMP Report 30.2R, Revised. New York: Columbia University Press, September 1945. pp. vii, 17; iv, 80; v, 57; iii, 25; iii, 18; iii, 39; ii, 41. $6.25. (London: Oxford University Press, 1946.)

REVIEWED BY JOHN W. TUKEY

Princeton University

Many of the features of this compendium are familiar to most of the readers of this review, but for the benefit of the others I shall enumerate them briefly. It consists of a heavy looseleaf binder containing 7 booklets of distinctive colors— each saddle stitched and usable separately. It is the last word (to date) in presenting sequential analysis to the statistician who may wish to use it in practice. It covers five elementary cases (each in a booklet, the two others being used for introduction and appendices):

> Acceptance or rejection by percent defective (Sec. 2)
> Comparative percent satisfactory (Double dichotomy) (Sec. 3)
> Acceptance or rejection by the adequacy of the mean (with known variability) (Sec. 4)
> Acceptance or rejection by the exact value of the mean (with known variability) (Sec. 5)
> Acceptance or rejection by the smallness of the variability (Sec. 6)

These cases are covered in complete detail, with illustrative examples, tables and charts. A copy should be accessible to every teacher of statistics and to every statistician in industry or experimental work who can propose new techniques of testing.

With this general introduction let us go on and explain what the reader will not find and what further work in this line the reviewer awaits with keen interest. The classical testing procedure was to test a sample of predetermined size and then decide to accept or reject. Long ago curtailed sampling and double sampling were developed to cut corners legitimately and reduce inspection costs. There are two situations, each more frequent in war than in peace, where it is clearly desirable to reduce the *average* number of items tested to a minimum:

(I) Where essentially all lots are accepted and the test is destructive so that the items tested are the main loss of production, or

(II) Where the cost of testing an item is large in comparison with the cost of production.

Subject to a practically unimportant allowance for the finite size of the lots, and to an allowance of unknown importance for the quality of lots presented, the methods of sequential analysis minimize this average number among all methods so far considered. When situation (I) or (II) holds without modifying complications, then, the best known method is sequential analysis, the natural descendant of double sampling. Otherwise, the situation is far from clear, and much judgement is involved in setting up a practically efficient scheme. The reader will get no help on this problem of judgement, nor in the problem of setting risks from the book under review—he will get every needed help with the mathematical problem of setting up a sequential plan to meet chosen risks, including complete tables of all necessary functions, including natural logarithms.

There is no reason to suppose that sequential analysis is the last word in testing procedures for the general problem of efficient testing, but what should be the next step ahead is not a step for the mathematical statistician. What is needed now is a careful analysis, by the operational research techniques so useful during the war, of a half-dozen industrial testing situations to determine what properties of the testing procedure are involved in cost and to what extent. Do we want the minimum average sample size, the minimum average square of the sample size—or what? With this there should go a corresponding operational study of the advantages of different OC curves, including those of what now seems to be a peculiar shape. Given these studies, we could put the problem in mathematical statistics to the mathematical statistician which he would then solve. But with the present lack of operational research groups in industry, it is probable that we will proceed in an unnatural way, and that the mathematical statistician will take the next step forward. For reasons of mathematical simplicity it is not unlikely that the sample plan with the minimum average squared sample size will come next.

The credit for the book is clearly assigned on the inside cover of each pamphlet in the following words: "So many members of the Statistical Research Group (Columbia) have participated in the preparation of this report, a previous edition of which was prepared by H. A. Freeman, that its authorship is attributed to the group as a whole. The responsibility for planning and preparing this

edition has been shared by H. A. Freeman, M. A. Girshick, and W. Allen Wallis, with the cooperation of Kenneth J. Arnold, Milton Friedman, Edward Paulson, and others. The theory of sequential analysis is mainly the work of A. Wald."

It may be of interest to notice a few minor points for the record. On page 1.01 it is indicated that 100% inspection is 100% effective— this seems far from industrial experience. Another badly needed set of operational studies would be on the influence of the sampling plan on inspector's inspection. On page 2.27, the footnote suggests that when a tabular procedure is used instead of a graphic one, that more decimal places should be kept—the logic of this is not clear. On page 4.14 it is stated that "similarly, if all patches had tested 400 minutes, the experiment would have terminated at 9.4 . . .". Clearly no such experiment can terminate after a fractional number of tests. On page A.09 it is stated that "Finally it should be mentioned that truncation of any kind ought generally to be avoided". This seems to the reviewer to be a rash statement, for when not only average sample size but all other properties entering into the practical efficiency of a sampling plan are considered, this decision will almost certainly be reversed. The relatively small number of these detailed points is an evidence of careful and competent workmanship.

A footnote to the Appendix (B) on some principles of sequential analysis states: "Any mathematician who may stray into this Appendix should be assured that the validity of the conclusions in no case depends upon the type of reasoning presented here; indeed, even for intuitive or heuristic arguments mathematicians may prefer those given in SRG 75". This warning and caveat seems unduly strong—the appendix is recommended to all mathematically minded newcomers to sequential analysis.

The same appendix warns the reader in a few places that the theory set forth does not allow for the fact that samples come in units. If the reader tries to apply the theory to cases far from normal inspection practice, for example with risks of 0.25 and average sample sizes of 12, he will then find out that this does occasionally make a difference. In conventional circumstances the approximation will not bother him.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Kurt W. Back has accepted a position with the Research Center for Group Dynamics, Massachusetts Institute of Technology.

Mr. Stanley D. Canter was discharged from the Army in October and has been enrolled as a graduate student in mathematical statistics at Columbia University.

Mr. William W. Cooper has accepted a position at Carnegie Institute of Technology, Pittsburgh.

Mr. Robert Dorfman is enrolled as a graduate student in the Department of Economics, University of California, Berkeley, and is also serving as a teaching assistant in that department.

Dr. Nicholas Fattu, formerly at Michigan State College, has accepted a teaching position at Indiana University, Bloomington.

Mr. John P. Gill is now Chief of the Research and Progress Analysis Division, War Assets Administration, Houston Regional Office, Texas.

Dr. Clausin D. Hadley has accepted a position with the Graduate School of Business, Stanford University.

Mr. Malcolm H. Henry is now Assistant Statistician in the Statistical Department of the Michigan State Department of Social Welfare, Lansing.

Dr. Alston S. Householder has accepted a position as Principal Physicist with the Monsanto Chemical Company, Clinton Laboratories, Oak Ridge, Tennessee.

Mr. Morton Kramer is now with the Office of International Health Relations, U. S. Public Health Service, Washington.

Mr. F. C. Leone, who was discharged from the military service in the fall, has returned to his former posisiton in the Department of Mathematics at Purdue University, Lafayette, Indiana.

Mr. Philip J. McCarthy, formerly at Princeton University, is now at Cornell University, Ithaca, New York.

Mr. Edward C. Molina has been named special lecturer in Mathematics at Newark College of Engineering, in addition to Dr. Emil J. Gumbel, previously mentioned.

Mr. Nicholas Pastore has accepted a position in the Department of Mathematics, City College of New York.

Dr. William S. Robinson is now Assistant Professor of Sociology and Statistics, University of California at Los Angeles.

Dr. Leonard J. Savage, who has a Special Rockefeller Fellowship, is spending the academic year at the Institute of Radiobiology and Biophysics, University of Chicago.

Professor Dunham Jackson died at Minneapolis on November 6, 1946. From 1919 until 1946 Mr. Jackson was Professor of Mathematics at the University of Minnesota, and in 1946 was named Professor Emeritus.

Professor Charles C. Wagner died suddenly on May 23, 1946, at the age of 52. He was acting dean of the College of Liberal Arts of Pennsylvania State College when he died.

Those interested in the work of the Mathematical Tables Project, will, upon request, be placed on the mailing list for copies of the monthly progress reports, issued by the Project. Requests should be addressed to Dr. Arnold N. Lowan, 150 Nassau St., New York, N. Y.

Statistical Research Laboratory, University of Michigan

Several developments in instruction and research in the general field of statistics are in progress at the University of Michigan.

At the beginning of the current academic year the new Statistical Research Laboratory was opened. It is planned that this unit, which is a division of the Graduate School, will serve as the center for research employing statistical methods and for research in statistical methodology. Free consultation and advice on statistical matters are offered to all members of the University engaged in research and the latest types of computing machines are available for their use at no cost to them. Or the Laboratory will undertake, at fees to cover costs, computing and the analysis of data for such individuals or units of the University. The Laboratory will have available the services of the University's completely equipped Sorting and Tabulating Station and expects to continue to provide a center for the most efficient computing service as improved machines are developed. The technical assistants employed by the Laboratory will be advanced students of statistics who will thus have the opportunity to supplement their training with experience with actual statistical investigations. Professor C. C. Craig as Director and Professor P. S. Dwyer are in charge of the new laboratory, each on a half-time basis.

The new Laboratory is a research and not a teaching unit and is distinct from the large statistical laboratories for the use of students in statistics courses already in existence on the campus. With respect to instruction in theoretical statistics, the curriculum in that subject in the Mathematics Department has recently been revised and extended to include twenty-four semester hours at the undergraduate and graduate levels in addition to courses in probability, finite differences, graphical methods, and quality control. The somewhat related professional program in actuarial mathematics has likewise been strengthened. The teaching staff for these two curricula includes Professors H. C. Carver, A. H. Copeland, C. C. Craig, P. S. Dwyer, C. H. Fischer, and C. J. Nesbitt.

A number of postwar research programs whose pursuit involves the use of probability and statistical methods have been established at the University of Michigan. Of especial interest is the new Survey Research Center under the

leadership of Professors R. Likert and A. A. Campbell who will continue activities begun by their group in Washington in the Department of Agriculture. Research by survey methods in the social sciences for public and private agencies and in survey methods themselves will be pursued and in addition a training program combining formal courses and apprenticeship in the Center is being set up.

New Members

The following persons have been elected to membership in the Institute:

Albert, George E., Ph.D. (Wisconsin) Head, Mathematics Division, Research Dept., Naval Ordnance Plant, Indianapolis, Ind., *1104 N. Oakland Ave.*

Ament, Richard P., B.A. (Cornell) Scientific Aid, *2129 20th St., N., Arlington, Va.*

Bennett, Myra S., (Mrs. C. A.). A.B. (Michigan) Office Mgr., Institute of Math. Stat., Rackham Bldg., Ann Arbor, Mich., *P. O. Box #3, Saline.*

Blankmeyer, Edith., A.B. (Western College) Stat., Res. Dept., National Broadcasting Co., 30 Rockefeller Plaza, New York 20, N. Y.

Blyth, Colin, Jr., M.A. (Queen's Univ. and Univ. of Toronto) Graduate student, Univ. of N. Car., Chapel Hill, N. Car., *209 Mangum Dormitory.*

Brown, Philip, B.S. (Pittsburgh) Stat., R. 329 Standard Oil Bldg., 3rd and Constitution Aves., Washington, D. C.

Bruno, O. P., B.M.E. (New York Univ.) Chief, Methods Section, Ballistic Research Labs., Aberdeen Proving Ground, Md.

Carrier, Norman H., M.A. (Cantab) Civil Servant, Mathematical Statistics Section of Chief Scientific Advisers Division, Ministry of Works, *c/o Westminster Bank, Palmers' Green, N. 13, London, England.*

Chand, Uttam, M.A. (Punjab Univ., India) Graduate student, Univ. of N. Car., Chapel Hill, N. Car., *112 Mangum Dormitory.*

Crow, Edwin L., Ph.D. (Wisconsin) Mathematician, Science Dept., Res., Devel., and Test Organization, USNOTS, Inyokern, Calif.

Dang, Mary., M.A. (California) Graduate student, Columbia University, New York 27, N. Y., *Box 267, Johnson Hall.*

Ens, Catherine C., B.S. (Dayton) Stat. Res. Ass't, Graduate School, Ohio State University, Columbus, Ohio, *267 Fifteenth Ave., Columbus 10.*

Fox, William H., Ph.D. (Indiana) Ass't Prof. of Educ. and Ass't Director of Res. and Field Service, Indiana Univ., Bloomington, Ind., *729 E. Hunter.*

Geisler, Murray A., M.A. (Columbia) Operations Analyst, Headquarters Army Air Forces, *222 N. Piedmont St., Arlington, Va.*

Gershenson, Charles P., B.B.A. (C.C.N.Y.) Res. Assoc., Institute of Psychological Res., *Box 130, Teachers College, New York 27, N. Y.*

Gilford, Leon, A.B. (Brooklyn) Econ. Analyst, Census Bureau, Washington, D. C., *1410 19th St., S. E.*

Goudsmit, S., A., Ph.D. (Leyden) Prof. of Physics, Northwestern Univ., Evanston, Ill.

Halperin, Max, M.S. (Iowa) Graduate student, Univ. of N. Car., Chapel Hill, N. Car., *211 No. Columbia.*

Halperin, Sidney L., Ph.D. (Ohio State) Psychologist, Neuropsychiatric Institute, Univ. of Mich. Hospital, Ann Arbor, Mich., *2401 Pittsfield Blvd., Pittsfield Village.*

Herbach, Leon H., A.B. (Brooklyn) Sub. Instr., Dept. of Math., Brooklyn Coll., N. Y., *1926 64th St., Brooklyn 4.*

Hoeffding, Wassily, Ph.D. (Berlin) 151 West 88 St., New York 24, N. Y.

Huhndorff, Roland F., B.S. (St. Mary's Univ.) Ass't to Ass't Chief Chemist, The Texas Co., Res. Lab., Port Arthur, Texas.

James, William C., A.B. (Knox Coll.) Director, Stat. Div., National Safety Council, 20 N. Wacker Dr., Chicago 6, Ill., *7835 So. Dobson Ave., Chicago 19.*

Lev, Joseph, Ph.D. (Cornell) Ass't Civil Service Examiner, N. Y. C. Civil Service Comm., and Lecturer, Teachers College, Columbia Univ., N. Y., *8550 Forest Parkway, Woodhaven 21.*

Linder, Arthur, Ph.D. (Bern) Prof. of applied math. stat., University of Geneva, Switzerland, *Avenue de Champel 24.*

Lord, Frederic M., M.A. (Minnesota) Ass't Director, Graduate Record Examination, 437 West 59th St., New York 19, N. Y., *153 W. 63rd St.*

Marshall, Herbert, B.A. (Toronto) Dominion Stat., Dominion Bureau of Statistics, , Ottawa, Canada.

Meacham, Alan D., Supv., Sorting and Tabulating Station, and Lecturer, School of Bus. Adm., Univ. of Mich., Ann Arbor, Mich., *114 Rackham Bldg.*

Miller, Irving, B.S. (C.C.N.Y.) Stat., Bureau of Labor Stat., Washington, D. C., *1900 Biltmore St., N. W., Washington 9.*

Nanda, D., N., M.A. (Agra, India) Graduate student, Univ. of N. Car., Chapel Hill, N. Car., *Dept. of Statistics.*

Pines, Sylvia F., M.A. (Michigan) Instr., Math. and Stat., *43-17 48th St., Long Island City 4, N. Y.*

Quastler, Henry M.D. (Vienna) Medical Radiologist, Carle Hospital Clinic, Urbana, Ill., *612 W. Nevada.*

Reiersol, Olav, Ph.D. (Stockholm) Teacher of stat., Univ. of Oslo, Oslo, Norway, *International House, 500 Riverside Dr., New York 27, N. Y.*

Romanovsky, Vsevolod I., Ph.D. (Moscow) Prof. at the Univ. and Member of the Academy of Sciences, Tashkend, U. S. S. R.

Rust, Charles H., S.J., M.A. (St. Louis) Graduate student, St. Louis Univ., St. Louis, Mo., *221 N. Grand Blvd., St. Louis 3.*

Seal, Hilary L., B.Sc. (Univ. Coll., London) Head of Stat. Branch, Room 2, Old Bldg., G., Admiralty, Whitehall, London, S. W. 1, England.

Serbein, Oscar N., Jr., M.S. (Iowa) Graduate student, Columbia Univ., New York 27, N. Y., *Army Hall, Rm. 333H, 1560 Amsterdam Ave., New York 31.*

Sholl, D., A., B.Sc. (London) Stat. in Math. Stat. Section of Chief Scientific Adviser's Div., Ministry of Works, *81 Lynmouth Ave., Bush Hill Park, Enfield, Middlesex, England.*

Siegel, Irving H., M.A. (New York) Chief, Economics Div., Veterans Adm., Washington, D. C., *5407 9th St., N. W., Washington 11.*

Sitgreaves, Rosedith, M.A. (Geo. Washington) Ass't Stat., U. S. Public Health Service (on leave); Graduate student, Columbia Univ., New York 27, N. Y., *Johnson Hall, 411 W. 116th St.*

Tama, Joseph, B.A. (Washington) Pfc. U. S. Army, 5250 TIC; GHQ AFPAC; APO 500, c/o Postmaster, San Francisco, Calif.

Tate, Merle W., Ed.M. (Harvard), M.A. (Montana) Assoc. Prof. of Educ., Hamilton Coll., Clinton, N. Y.

Thrall, Robert M., Ph.D. (Illinois) Ass't Prof. of Math., Univ. of Mich., Ann Arbor, Mich., *953 Spring St.*

Vaughn, Kenneth W., Ph.D. (Iowa) Director, Graduate Record Examination Office of the Carnegie Foundation for the Advancement of Teaching; and, Assoc. Director of Co-operative Test Service of Amer. Council on Educ., *437 West 59th St., New York 19, N. Y.*

Wallace, Clifford A., Sup't of Quality, Camera Works, Eastman Kodak Co., 333 State St., Rochester, N. Y.

Wilkins, J., Ernest, Jr., Ph.D. (Chicago) Mathematician, American Optical Co., S. I. D., Box A, Buffalo 15, N. Y.

Wilkinson, Roger I., B.S.E.E. (Iowa State) Member Technical Staff, Bell Telephone Labs., 463 West St., New York, N. Y.

REPORT ON THE BOSTON MEETING OF THE INSTITUTE

The twenty-fourth meeting of the Institute of Mathematical Statistics was held at the Hotel Statler, Boston, Massachusetts, on Saturday, December 28, 1946. The meeting was held in conjunction with the One Hundred Thirteenth Annual Meeting of the American Association for the Advancement of Science. The following 45 members of the Institute attended the meeting:

K. J. Arnold, M. S. Bartlett, W. D. Baten, C. I. Bliss, G. W. Brier, G. W. Brown, T. H. Brown, B. H. Camp, C. W. Churchman, W. G. Cochran, J. H. Curtiss, D. B. DeLury, P. V. Dorweiler, Churchill Eisenhart, Benjamin Epstein, H. A. Freeman, Hilda Geiringer, H. H. Germond, J. A. Greenwood, Boyd Harshbarger, W. A. Hendricks, E. H. C. Hildebrandt, W. C. Jacob, H. B. Kaitz, L. F. Knudsen, Walter Leighton, A. J. Lotka, J. W. Mauchly, Margaret Merrell, E. B. Mode, Frederick Mosteller, C. M. Mottley, Doris Newman, R. H. Noel, H. W. Norton, Otis Pope, C. J. Rees, C. F. Roos, P. J. Rulon, J. W. Tukey, W. M. Upholt, F. M. Wadley, C. L. Weaver, C. P. Winsor, W. J. Youden.

At the morning session, a joint session with the Biometrics Section of the American Statistical Association, the following program was presented with Professor E. B. Wilson of Harvard University as chairman:

Topic: *The Analysis of Variance in Biology*

Papers: *The Assumptions Underlying the Analysis of Variance*
Professor Churchill Eisenhart, University of Wisconsin and The National Bureau of Standards
Some Consequences when the Assumptions are not Satisfied
Professor W. G. Cochran, North Carolina State College
The Use of Transformations
Professor M. S. Bartlett, Cambridge University and the University of North Carolina

Discussion: Professor Boyd Harshbarger, Virginia Polytechnic Institute
Dr. W. C. Jacob, Long Island Vegetable Research Farm
Professor C. P. Winsor, Johns Hopkins University
Dr. W. J. Youden, Boyce Thompson Institute

The program for the afternoon session, also a joint session with the Biometrics Section, under the chairmanship of Dr. E. J. DeBeer, Wellcome Research Laboratories, was as follows:

Topic: *The Analysis of Variance in Biology* (*continued*)

Papers: *The Analysis of Covariance*
Professor D. B. DeLury, Virginia Polytechnic Institute
Discriminant Functions
Professor George W. Brown, Iowa State College

Discussion: Professor W. D. Baten, Michigan State College
Professor C. I. Bliss, Yale University
Mr. W. A. Hendricks, U. S. Department of Agriculture

P. S. Dwyer,
Secretary.

149

ANNUAL REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1946

New Opportunities

The return to peacetime conditions presents the Institute with new opportunities for expanding its activities and usefulness. An increased appreciation for mathematical statistics has followed the many contributions made by our members to the war effort. The numerous societies interested in specific applications of statistics have come to look to the Institute both for leadership in theory and for playing its part in the dissemination of new results. As a result of the drastic interruption in the normal training of students during the war, there is unusually keen competition for the services of capable statisticians. Those of our members who are engaged in teaching are responsible for the execution of a vigorous training program to meet current and future demands promptly and without sacrifice of quality. In short, we are in a position, as never before, to advance the development and efficient use of mathematical statistics. The following account of some of our activities during the year will indicate, I believe, that the record is creditable. Yet in many instances what has been accomplished is only a beginning.

Meetings

The Development Committee has repeatedly stressed the desirability of an extension in our customary schedule of meetings in order to provide additional contacts between mathematical statisticians and the users of statistics. Owing to the greater availability of railway and hotel accommodation in 1946, we obtained our first opportunity to put this extension into effect. The regular winter meeting with the American Statistical Association and other social science organizations was resumed at Cleveland in January, while the late summer meeting with the mathematicians took place at Cornell in September. In addition, two meetings were held with different sections of the American Association for the Advancement of Science, at St. Louis in March and at Boston in December. On both occasions the programs were expository and attracted large audiences. Finally, at the invitation of Princeton University, a one-day meeting at Princeton in November was devoted to the analysis of variance. While no joint sessions were conducted with engineering or industrial societies, several of our members took prominent parts in the programs of such societies.

For the near future, it seems desirable to continue the practice of meeting in the winter with the ASA and social science groups and in the summer with the mathematical groups. In 1947 these meetings will be at Atlantic City, January 24–27 and at Yale, September 1–5 respectively. It is not known whether conditions in future years will produce a return to Christmas rather than January meetings: for the present the hotel situation swings the balance in favor of January.

In 1946 the membership of the program committee was enlarged so that it would be better equipped to arrange joint meetings with other societies. We owe our thanks to the members for their successful efforts in the face of difficulties which still attend the planning of a meeting.

ANNALS

Despite the scarcity of manuscripts in the later stages of the war, our editor, Professor S. S. Wilks, succeeded throughout in maintaining the annual volumes of the Annals at their usual size. During 1946, scarcity gave way to plenty. The number of papers of good quality submitted in recent months is sufficiently great that there will be more than enough, by current estimates, to fill the 1947 volume. To narrow the scope of the Annals or to reject good papers would be undesirable. Accordingly, the Directors have authorized an increase of 100 pages in the 1947 volume if this is necessary to insure the publication of all acceptable papers.

A gratifying testimony to the prominence of the Annals in its field is the marked increase in the demand for back numbers. Our Secretary-Treasurer reports that sales amounted to $3,235. To meet actual or anticipated orders, eleven issues were reprinted during 1946 at a cost of $2,809.

For most members of the Institute, even those who serve on the Board, work on Institute affairs occupies only a minor portion of our time. The editor is never free from some forthcoming publication deadline. Initial perusal of manuscripts, selection of referees, editorial decisions, handling of the production phases of publication and much miscellaneous correspondence (not all of it pleasant) make editorial work a daily preoccupation, year in and year out. An annual word of thanks is an inadequate expression of our indebtedness to Professor Wilks.

MEMBERSHIP AND FINANCE

At the beginning of 1945 there were 606 members. A year later this figure had increased to 777 and at the end of 1946 the figure stood at 900. A fifty percent increase in two years is another evidence of the healthy growth of the Institute. It has been attained to a considerable extent through the hard work of our Secretary-Treasurer, P. S. Dwyer and the cooperation of individual members.

The Secretary-Treasurer also reports a very satisfactory net gain in assets of $2,627 during the year. Nevertheless, financial problems may arise in the near future. Printing and other costs have risen sharply, and the printing of an enlarged Annals will be an additional drain on our resources. Both the Membership and Development Committees have given some thought to the need for additional revenue that may face us soon. They have recommended consideration of the possibility of Institutional Memberships, a device that has been found satisfactory by some other societies. A continued growth in membership will also help greatly to finance expanded activities.

COMMITTEES

Inter-society affairs: The report of the 1944 Committee on Development, stressing the need for closer cooperation amongst the various societies interested in statistics, provided the stimulus for active efforts in this direction. A meeting of representatives of these societies was called early in 1945 at the invitation of the American Statistical Association. This meeting suggested that a reconstitution of the ASA might enable it to become the central binding organization. Accordingly, a committee of the ASA has worked for a considerable time on a revision of the ASA constitution, which it is intended to submit to the votes of ASA members early in 1947. The new constitution provides for representation from other societies on the Council of the ASA, should these societies decide to associate or affiliate with the ASA.   .

From our own point of view, it has seemed wise to delay action on certain internal affairs while awaiting the outcome of these developments in the ASA. Thus a statement of policy with regard to the formation of chapters of the IMS is needed and the problem has been considered both by a special committee in 1945 and by the Development Committee in 1946. The latter committee recommends that no decision be made pending examination of the provisions for joint sponsorship of local and regional chapters in the new ASA constitution. Similarly, our own Committee on Revising the Constitution and By-Laws has deferred a final report until the attitude of our members towards the new developments can be expressed. It is to be hoped that decisions can be taken in 1947.

Tabulation: The advances made in recent years in the construction of new types of computing equipment justified an enlargement of our Committee on Tabulation, which now includes experts both on the building of machines and on the calculation and use of tables. The committee plans to keep our members informed of progress in this field.

Government Service: Dr. W. Edwards Deming served as chairman of a new committee on Mathematical Statistics and Statisticians in the Government Service. Although the federal government employs many mathematical statisticians, explicit recognition of the profession is lacking in many instances. As has happened in other fields, statisticians are sometimes officially classed as economists and little provision is made for mathematical statisticians in recruitment policies. Moreover, it is probable that a number of branches of the government, at present unaware of the functions of a statistician, could employ several with profit. The new committee will endeavor to insure that mathematical statistics is recognized and effectively utilized in the federal service.

Assistance to libraries: Like other professional societies, the Institute has received a number of appeals from libraries in war areas whose periodicals were looted or destroyed during the war. After careful consideration, the Board decided that official action should be limited to the free provision of missing copies of the Annals to all former subscribers who intend to renew subscriptions for the future. In addition, a committee with Professor J. Neyman as chairman

was appointed to establish a procedure by which gifts of individual members (books, reprints, back numbers of the Annals or cash for the purchase of back numbers) could be handled. At the suggestion of this committee a general appeal for the small sum of 50 cents per member was circulated with the December billing. Individual collections are also being made at certain centers.

Teaching: The Committee on Teaching has not made as much progress as it would have liked, owing to the dispersal of its members and the taking up of new civilian posts. Members have, however, cooperated with the Committee on Applied Mathematical Statistics of the National Research Council, which is engaged on a somewhat similar survey.

Rietz lecture: The first lecturer in the new series of lectures in honor of the late Henry Lewis Rietz will be Professor A. Wald. His topic will be "Sequential Estimation and Multi-Decisions". The lecture will be delivered in connection with the Yale meetings, September 1947.

Representatives: In addition to its committee work, the Institute cooperates, through representatives, with the Division of Physical Sciences of the National Research Council, the Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing, the American Association for the Advancement of Science, the Inter-Society Committee on Federation and the Policy Committee for Mathematics. The last committee, which was appointed in 1946, will consider important problems that affect the mathematics profession as a whole.

Nominations: The Committee on Nominations, consisting of Professor P. R. Rider (chairman), Professor B. H. Camp and Professor G. M. Cox, has made the following nominations for officers in 1947.

President:	W. Feller
Vice-Presidents:	J. H. Curtiss
	M. H. Hansen
Secretary-Treasurer:	P. S. Dwyer

While it is perhaps improper to comment on nominations, I should like to express my personal appreciation of Professor Dwyer's action in being willing to offer himself for re-nomination as Secretary-Treasurer. The successful operation of the Institute rests mainly on the Secretary-Treasurer, and the demands of the Office are even more continuous and exacting than those on the editor. Professor Dwyer's splendid work during his first three years of office, carried on at considerable sacrifice of his research interests, deserves the best thanks and appreciation of every member.

In conclusion, it is a pleasure to express my sincerest thanks to all committee chairmen and members and to all representatives for their excellent work for the good of the Institute, and to all Institute members for their loyal support.

W. G. Cochran,
President, 1946.

Committees of the Institute

Committee	*Personnel*
Development	E. G. Olds (chairman), C. I. Bliss, M. A. Girshick, F. C. Mosteller, P. S. Olmstead, H. Scheffé.
Membership	W. Feller (chairman), C. C. Craig, P. A. Horst, T. Koopmans
Program	J. H. Curtiss (chairman), M. Friedman, B. Harshbarger, W. N. Hurwitz, A. M. Mood, F. C. Mosteller, J. W. Tukey
Mathematical Statistics and Statisticians in the Government Service	W. E. Deming (chairman)
Revising the Constitution- and By-Laws	M. H. Hansen (chairman), C. I. Bliss, A. T. Craig, J. H. Curtiss, W. Shewhart
Tabulation	C. Eisenhart (chairman), P. S. Dwyer, H. Goldstine, A. N. Lowan, H. W. Norton, G. R. Stibitz
Teaching	H. Hotelling (chairman), W. Bartky, W. E. Deming, M. Friedman
Nominations	P. R. Rider (chairman), B. H. Camp, G. M. Cox
Finance	P. S. Dwyer (chairman), L. A. Knowler, C. F. Roos, F. F. Stephan
Subscription to Purchase Annals for Countries Devastated by War	J. Neyman (chairman), W. Feller, P. L. Hsu

Society	*Representatives*
Inter-Society Committee on Federation	J. H. Curtiss, P. S. Olmstead
Policy Committee for Mathematics	W. Feller

Society	*Representatives*
Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing	F. C. Mosteller, S. S. Wilks
American Association for the Advancement of Science	G. W. Snedecor
Division of Physical Sciences, NRC	W. Bartky

REPORT OF THE SECRETARY-TREASURER OF
THE INSTITUTE FOR 1946

The Institute of Mathematical Statistics held five meetings during 1946, at Cleveland on January 24–27, at St. Louis on March 30, at Ithaca on August 22–23, at Princteon on November 1, and at Boston on December 28.

The large number of meetings has necessitated frequent mailings to the membership. Memoranda to members, with appropriate enclosures, were sent out in January, March, June, July, October, and November.

The Secretary-Treasurer wishes to acknowledge the cooperation of the members of the Institute in paying bills promptly, in considerable activity leading to an increase in membership, and in general looking after the interests of the Institute.

At the beginning of 1946 the Institute had 777 members. During the year 180 new members joined the Institute, an increase of 23%. However, during 1946 the Institute lost 57 members. Of these, 15 resigned, 37 were dropped for non-payment of dues, and 5 are deceased. Some of the 37 dropped we have been unable to contact, and it is very probable that, in some cases, membership will be resumed in the future. The net increase in members during the year was 123, or about 16%, making a total of 900 members.

The following members died during the year:

> Professor O. F. Banos
> Professor S. A. Cudmore
> Professor Dunham Jackson
> Dr. Walter F. Schilling
> Professor C. C. Wagner

The office of the Secretary-Treasurer sent a reprint of an *Annals* article and information about the Institute to 1800 persons interested in Quality Control. At least 28 of the new members became members as a result of this drive. As a continuation of a campaign started in 1945, the Institute also sent literature about the *Annals* to several hundred libraries and laboratories.

The Secretary-Treasurer wishes to acknowledge the continued assistance of Professor Lloyd Knowler in caring for the back issues of the *Annals* which are stored at Iowa City.

A few comments about the financial statement which appears below are in order. In addition to the increase in membership, mentioned above, the chief rise in income resulted from the unprecedented sales in back issues which amounted to $3,234.88, an increase over the preceding year (the previous high) of 86%. These heavy sales, however, depleted the supplies of many of our early issues, so that we were forced to reprint eleven of these issues and also the cumulative index during the year. This cost $2,809.00 (for 500 copies of each) and indicates that a much larger portion of our assets is in inventory, as shown in Exhibit D

Following the instructions of the Finance Committee, Professor H. C. Carver was paid for his share of all issues in which he and the Institute had joint ownership.

Nine members have paid life memberships during the year, increasing the total of life membership funds by $812.50.

The net gain in assets of $2,627.23 is very satisfactory even though this gain is evident in increased inventory and not in a better cash position.

FINANCIAL STATEMENT

December 31, 1945, to December 31, 1946

A RECEIPTS

BALANCE ON HAND, DECEMBER 31, 1945	$7,548 22
DUES	4,638.40
LIFE MEMBERSHIP PAYMENTS	812 50
SUBSCRIPTIONS	2,057 54
SALE OF BACK NUMBERS	3,234 88
INCOME FROM INVESTMENTS	150 00
MISCELLANEOUS	121 29
TOTAL	$18,562 83

B EXPENDITURES

ANNALS—CURRENT		
Office of Editor	$125 00	
Waverly Press	4,566 27	
		$4,691 27
ANNALS—BACK NUMBERS		
Purchase from H C. Carver	644 50	
Reprinted 500 copies	2,809 00	
Vol I #1, II #2, II #3, III #3, IV #1, VII #3, VII #2, VIII		
#1, 2, 3, 4, Cumulative Index		
Iowa City Office	41 46	
Binding	68 00	
		3,562 96
OFFICE OF PRESIDENT		25 62
MATHEMATICAL REVIEWS		100 00
OFFICE OF THE SECRETARY-TREASURER		
Printing, Mimeographing, programs, etc (including stamped envelopes)	$967 14	
Printing 1800 copies of Wald-Wolfowitz article	140 00	
Postage and supplies	375 00	
Clerical help	1,420 25	
		2,902.39
MISCELLANEOUS		39 04
BALANCE ON HAND, DECEMBER 31, 1946 (Cash and Bonds)		7,241.55
		$18,562 83

C. Summary of Receipts and Expenditures

Balance on Hand,* December 31, 1945	$7,548.22
Receipts during 1946	11,014.61
Expenditures during 1946	11,321.28
Balance on Hand,* December 31, 1946	7,241.55

D. Comparison of Assets on December 31, 1945 and December 31, 1946

	1945	1946
US Government G Bonds	$6,000 00	$5,000 00
Life Membership Funds	{ 888 00	{ 1888.00 Bonds
	{ 327.00	{ 139 50 Bank Dep.
Additional Bank Deposits	333.22	214.05
Current Accounts Receivable	255 35	452 62
Estimated Value (Cost) of back issues of *Annals*	4,497.95	7,234 58**
Total	$12,301.52	14,928.75
Net Gain 1946		2,627.23

E. Liabilities of Institute of Mathematical Statistics as of December 31, 1946

All bills which have been presented have been paid and there are no outstanding accounts against the Institute   The $2027.50 in Life Membership payments require the Institute to provide the privileges of membership for life for the 26 members who have made payments. Also, $2686 71 should be credited to 1947 dues and subscriptions

Paul S. Dwyer
Secretary-Treasurer.

December 31, 1946

---

* In form of bank deposit and government bonds.

** Value of *Annals* calculated at 67 cents per copy, and based on physical inventory!

ANNUAL REPORT OF THE EDITOR FOR 1946

During 1946 there was a considerable increase in the number of manuscripts submitted to the Editorial Committee of the *Annals*. A total of 49 papers including 18 short notes were published in the 1946 volume of the *Annals*. The publication of these papers together with various official reports of the Institute and the Directory of the Institute required a total of 555 pages. Plans are already under way to expand the 1947 volume of the *Annals* to 600 pages.

During recent years there has been a very noticeable broadening of interest in the field of probability and statistical theory on the part of readers and contributors to the *Annals*. Contributors to the 1946 volume came from university departments of astronomy, biology, mathematics, sociology and statistics; from Army, Navy and other government groups; and from industrial laboratories and quality control departments. More recently, contributions have been received from physicists, chemists and other groups. More contributions are now being received from overseas than in previous years. Every effort is being made to keep the *Annals* balanced with respect to these various directions of interest in probability and statistical problems. It is believed that one of the most effective things which could be done for the readers of the *Annals* is to publish expository articles from time to time on new fields of development in probability and statistical theory. Invitations have been accepted by several individuals to prepare such articles.

Dr. Thornton C. Fry has asked to be relieved from the Editorial Committee, as of January 1, 1947. The Editor wishes to take this opportunity to express his gratitude for the service which Dr. Fry has rendered in connection with the editorial work on the *Annals* during the past nine years.

On behalf of the Editorial Committee for the *Annals*, the Editor wishes to acknowledge with thanks the refereeing assistance which has been provided by the following persons during 1946: R. L. Anderson, T. W. Anderson, David Blackwell, Z. W. Birnbaum, K. L. Chung, W. J. Dixon, J. L. Doob, M. A. Girshick, T. E. Harris, L. Henkin, M. Kac, Irving Kaplansky, Bradford F. Kimball, T. Koopmans, H. Levene, H. B. Mann, P. J. McCarthy, F. C. Mosteller, H. E. Robbins, D. F. Votaw, J. E. Walsh and C. P. Winsor. The Editor is also indebted to the following individuals at Princeton University for preparation of manuscripts for the printer, and other editorial assistance: Mrs. Gladys B. Huling, Mrs. Eleanor C. Schoenly and J. E. Walsh.

<div align="right">

S. S. WILKS
Editor.

</div>

December 31, 1947

CONSTITUTION

OF THE

INSTITUTE OF MATHEMATICAL STATISTICS

ARTICLE I

NAME AND PURPOSE

1. This organization shall be known as the Institute of Mathematical Statistics.
2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

MEMBERSHIP

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.

2. Voting members of the Institute shall be (a) the Fellows, and (b) all others, Junior members excepted, who have been members for twenty-three months prior to the date of voting.

3. No person shall be a Junior Member of the Institute for more than a limited term as determined by the Committee on Membership and approved by the Board of Directors.

ARTICLE III

OFFICERS, BOARD OF DIRECTORS, AND COMMITTEE ON MEMBERSHIP

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer. The terms of office of the President and Vice-Presidents shall be one year and that of the Secretary-Treasurer three years. Elections shall be by majority ballots at Annual Meetings of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers, the two previous Presidents, and the Editor of the Official Journal of the Institute.

3. The Institute shall have a Committee on Membership composed of a Chairman and three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

ARTICLE IV

MEETINGS

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such

time as the Board of Directors may designate. Additional meetings may be called from time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute.

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the Board by the Secretary-Treasurer at least five days prior to the date set therefor. Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3. Meetings of the Committee on Membership may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting. Committee business may also be transacted by correspondence if that seems preferable.

4. At a regularly convened meeting of the Board of Directors, four members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

ARTICLE V

PUBLICATIONS

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. The Editor of the *Annals of Mathematical Statistics* shall be a Fellow appointed by the Board of Directors of the Institute. The term of office of the Editor may be terminated at the discretion of the Board of Directors.

2. Other publications may be originated by the Board of Directors as occasion arises.

ARTICLE VI

EXPULSION OR SUSPENSION

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

ARTICLE VII

AMENDMENTS

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

BY-LAWS

ARTICLE I

DUTIES OF THE OFFICERS, THE EDITOR, BOARD OF DIRECTORS, AND COMMITTEE ON MEMBERSHIP

1. The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present, shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be submitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2. The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute and once a year he shall publish in the *Annals of Mathematical Statistics* a classified list of all Members and Fellows of the Institute. He shall send out calls for annual dues and acknowledge receipt of same; pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditor shall report to the President.

3. Subject to the direction of the Board, the Editor shall be charged with the responsibility for all editorial matters concerning the editing of the *Annals of Mathematical Statistics*. He shall, with the advice and consent of the Board, appoint an Editorial Committee of not less than twelve members to co-operate with him; four for a period of five years, four for a period of three years, and the remaining members for a period of two years, appointments to be made annually as needed. All appointments to the Editorial Committee shall terminate with the appointment of a new Editor. The Editor shall serve as editorial adviser in the publication of all scientific monographs and pamphlets authorized by the Board.

4. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute. The power of election to the different grades of Membership, except the grades of Member and Junior Member, shall reside in the Board.

5. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the

different grades of membership. The Committee shall review these qualifications periodically and shall make such changes in these qualifications and make such recommendations with reference to the number of grades of membership as it deems advisable. The power to elect worthy applicants to the grades of Member and Junior Member shall reside in the Committee, which may delegate this power to the Secretary-Treasurer, subject to such reservations as the Committee considers appropriate. The Committee shall make recommendations to the Board of Directors with reference to placing members in other grades of membership. The Committee shall give its attention to the question of increasing the number of applicants for membership and shall advise the Secretary-Treasurer on plans for that purpose.

ARTICLE II

Dues

1. Members shall pay five dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members shall pay five dollars annual dues. The annual dues of Junior Members shall be two dollars and fifty cents.

The annual dues of Fellows shall be five dollars. The annual dues of Sustaining Members shall be fifty dollars. Honorary Members shall be exempt from all dues.

(a) Exception. In the case that two Members of the Intitute are husband and wife and they elect to receive between them only one copy of the Official Journal, the annual dues of each shall be three dollars and seventy-five cents.

(b) Exception. Any Member or Fellow may make a single payment which will be accepted by the Institute in place of all succeeding yearly dues and which will not otherwise alter his status as a Member or Fellow. The amount of this payment will depend upon the age of this Member or Fellow and will be based upon a suitable table and rate of interest, to be specified by the Board of Directors.

(c) Exception. Any Member or Junior Member of the Institute serving, except as a commissioned officer, in the Armed Forces of the United States or of one of its allies, may upon notification to the Secretary-Treasurer be excused from the payment of dues until the January first following his discharge from the Service. He shall have all privileges of membership except that he shall not receive the Official Journal. However during the first year of his resumed regular membership he may have the right to purchase, at $2.50 per volume, one copy of each volume of the Official Journal published during the period of his service membership.

2. Annual dues shall be payable on the first day of January of each year.

3. The annual dues of a Fellow, Member, or Junior Member include a subscription to the Official Journal. The annual dues of a Sustaining Member include two subscriptions to the Official Journal.

4. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues may be six months in arrears, and to accompany such notice by a copy of this Article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent one to the Board of Directors, by whom the person's name may be stricken from the rolls and all privileges of membership withdrawn. Such person may, however, be re-instated by the Board of Directors upon payment of the arrears of dues.

ARTICLE III

SALARIES

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee.

ARTICLE IV

AMENDMENTS

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors.

PROBLEMS IN PROBABILITY THEORY

By Harald Cramér

University of Stockholm

1. Introduction. The following survey of problems in probability theory has been written for the occasion of the Princeton Bicentennial Conference on "The Problems of Mathematics," Dec. 17–19, 1946. It is strictly confined to the purely mathematical aspects of the subject. Thus all questions concerned with the philosophical foundations of mathematical probability, or with its ever increasing fields of application, will be entirely left out.

No attempt to completeness has been made, and the choice of the problems considered is, of course, highly subjective. It is also necessary to point out explicitly that the literature of the war years has only recently—and still far from completely—been available in Sweden. Owing to this fact, it is almost unavoidable that this paper will be found incomplete in many respects.

I. FUNDAMENTAL NOTIONS

2. Probability distributions. From a purely mathematical point of view, probability theory may be regarded as the theory of certain classes of *additive set functions*, defined on spaces of more or less general types. The basic structure of the theory has been set out in a clear and concise way in the well-known treatise by Kolmogoroff [53]. We shall begin by recalling some of the main definitions. . Note that the word *additive*, when used in connection with sets or set functions, will always refer to a *finite or enumerable* sequence of sets.

Let ω denote a variable point in an entirely arbitrary space Ω, and consider an additive class C of sets in Ω, such that the whole space Ω itself is a member of C. Further, let $P(S)$ be an additive set function, defined for all sets S belonging to the class C, and suppose that

$$P(S) \geqq 0 \text{ for all } S \text{ in } C,$$
$$P(\Omega) = 1.$$

We shall then say that $P(S)$ is a *probability measure*, which defines a *probability distribution in* Ω. For any set S in C, the quantity $P(S)$ is called the *probability* of the *event* expressed by the relation $\omega \subset S$, i.e. the event that the variable point ω takes a value belonging to S. Accordingly we write

$$P(S) = P(\omega \subset S).$$

Suppose now that $\omega' = g(\omega)$ is a function of the variable point ω, defined throughout the space Ω, the values ω' being points of another arbitrary space Ω'. Let S' be a set in Ω' and denote by S the set of all points ω such that $\omega' = g(\omega)$ belongs to S'. Whenever S belongs to C, we define a set function $P'(S')$ by writing

$$P'(S') = P(S).$$

It is then easy to see that $P'(S')$ is defined for all S' belonging to a certain additive class C' in the new space Ω', and that $P'(S')$ is a probability measure in Ω', such that $P'(S')$ signifies the probability of the event $\omega' \subset S'$ (which is equivalent to $\omega \subset S$). We shall say that $P'(S')$ is attached to the probability distribution in Ω' which is *induced* by the given distribution in Ω and the function $\omega' = g(\omega)$.

3. Random variables. Consider in particular the case when ω' is a real number ξ, such that $\xi = g(\omega)$ is a real-valued C-measurable function of the argument ω. Then C' includes the class B_1 of all Borel sets S' of the space $\Omega' = R_1$ of all real numbers, and we shall call ξ a *one-dimensional real random variable*. The probability of the event $\xi \subset S'$ is uniquely defined for any Borel set S' of R_1, as soon as the function

$$F(x) = P(\xi \leqq x)$$

is known for all real x. $F(x)$ is called the *distribution function* $(d.f.)$ of the random variable ξ. If the function $\xi = g(\omega)$ is integrable over Ω with respect to the measure $P(S)$, we write

$$\xi = \int_\Omega g(\omega)\ dP = \int_{-\infty}^\infty x\ dF(x),$$

and denote this expression as the *expectation* or *mean value* of the random variable ξ. Any real-valued B-measurable function $\eta = h(\xi)$ is also a random variable with the probability distribution induced by the original ω-distribution and the function $\eta = h(g(\omega))$. If η is integrable over Ω with respect to P, its mean value may be written in the form

$$E\eta = Eh(\xi) = \int_\Omega h(g(\omega))\ dP = \int_{-\infty}^\infty h(x)\ dF(x).$$

More generally, if $\omega' = (\xi_1, \cdots, \xi_n)$ is a point in an n-dimensional Euclidean space R_n, while C' includes the class B_n of all Borel sets of R_n, we are concerned with an *n-dimensional real random variable*. The distribution of this variable, which is also called the joint distribution of the n one-dimensional variables $\xi_1, \cdots, \xi_n$, is uniquely defined, as soon as the joint d.f.

$$F(x_1, \cdots, x_n) = P(\xi_1 \leqq x_1, \cdots, \xi_n \leqq x_n)$$

is known for all real $x_1, \cdots, x_n$.

The variables $\xi_1, \cdots, \xi_n$ are said to be *independent*, if $F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$, where $F_\nu(x_\nu)$ is the d.f. of the variable ξ_ν.

The extension to *complex random variables* is obvious. Suppose e.g. that $\xi = g(\omega)$ and $\eta = h(\omega)$ are two one-dimensional real variables, and consider the complex variable $\xi + i\eta = g(\omega) + ih(\omega)$. By definition, we identify the distribution of this variable with that of the two-dimensional real variable (ξ, η), and we put

$$E(\xi + i\eta) = E\xi + iE\eta.$$

Joint distributions of several complex variables are introduced in a corresponding way.

4. Characteristic functions. If ξ is a one-dimensional real random variable, the mean value

$$\varphi(z) = Ee^{iz\xi} = \int_{-\infty}^{\infty} e^{izx} \, dF(x)$$

exists for all real z, and we have

$$|\varphi(z)| \leq 1, \qquad \varphi(0) = 1.$$

$\varphi(z)$ is called the *characteristic function* (*c.f.*) of the distribution corresponding to the variable ξ. The reciprocal formula (Lévy)

$$F(x) - F(y) = -\frac{1}{2\pi i} \lim_{z \to \infty} \int_{-z}^{z} \frac{e^{-izx} - e^{-izy}}{z} \varphi(z) \, dz,$$

which holds for any continuity points x and y of F, shows that there is a one-one correspondence between the d.f. $F(x)$ and the c.f. $\varphi(z)$. As we shall see below, the c.f. provides a powerful analytical tool for operations with probability distributions.

When a complex-valued function $\varphi(z)$ of the real variable z is given, it is often important to be able to decide whether $\varphi(z)$ is or is not the c.f. of some distribution. If we assume a priori that $\varphi(0) = 1$, each of the following conditions is necessary and sufficient for $\varphi(z)$ to be a c.f.

$\cdot$.*A*. $\varphi(z)$ should be bounded and continuous for all z, and such that the integral

$$\int_{0}^{A} \int_{0}^{A} \varphi(z - u)e^{ix(z-u)} \, dz \, du$$

is real and non-negative for all real x and all $A > 0$ (Cramér [11], in simplification of an earlier result due to Bochner, [4]).

B. There should exist a sequence of functions $\psi_1(z), \psi_2(z), \cdots$ such that

$$\varphi(z) = \lim_{n \to \infty} \int_{-\infty}^{\infty} \psi_n(x + z)\overline{\psi_n(x)} \, dx$$

holds uniformly in every finite z-interval (Khintchine, [45]).

These general theorems are not always easy to apply in practice. Among less general results which are more easily applicable, we mention the almost trivial fact that a function $\varphi(z)$ which near $z = 0$ is of the form $\varphi(z) = 1 + o(z^2)$ cannot be a c.f. unless $\varphi(z) = 1$ for all z, and the two following theorems:

1) An integral function $\varphi(z)$ of order $\gamma < 1$ can never be a c.f. (Lévy, [64]), and

2) an integral function $\varphi(z)$ of finite order $\gamma > 2$ cannot be a c.f. unless the convergence exponent of its zeros is equal to γ (Marcinkiewicz, [72]). The latter result shows e.g. that no function of the form $e^{g(z)}$, where $g(z)$ is a polynomial of degree > 2, can be a c.f.

It would be highly desirable to obtain further results in this direction.

The c.f. of the joint distribution of n real random variables $\xi_1, \cdots, \xi_n$ is the function $\varphi(z_1, \cdots, z_n)$ defined by the relation

$$\varphi(z_1, \cdots, z_n) = Ee^{i(z_1\xi_1 + \cdots + z_n\xi_n)}.$$

Most of the above results for c.f. in one variable can be directly generalized to the multi-variable case.

5. Random sequences and random functions. Let t be a variable point in an arbitrary space **T**, and consider the space Ω, where each point ω is a real-valued function $\omega = x(t)$ of the variable argument t. Let $t_1, \cdots, t_n$ be any finite set of distinct points t. The set of all functions $\omega = x(t)$ satisfying the inequalities

$$a_j < x(t_j) \le b_j, \, (j = 1, \cdots, n),$$

will be called an *interval* in the space Ω. The Borel sets in Ω will be defined as the smallest additive class B of sets in Ω containing all intervals.

Suppose now that, for any choice of n and the t_j, the variables $x(t_1), \cdots, x(t_n)$ are random variables having a known n-dimensional joint distribution. If the family of all distributions corresponding in this way to finite sequences $t_1, \cdots, t_n$ satisfies certain obvious consistency conditions, a fundamental theorem due to Kolmogoroff asserts that this family determines a unique probability distribution in the space Ω of all functions $x(t)$. The corresponding probability

$$P(S) = P(x(t) \subset S)$$

is uniquely defined for all Borel sets S of Ω.

Consider in particular the case where **T** is the set of non-negative integers $t = 0, 1, 2, \cdots$. The space Ω then is the space of all sequences $(x_0, x_1, \cdots)$ of real numbers. As soon as the joint distribution of any finite number of variables $x_{\nu_1}, \cdots, x_{\nu_n}$ is defined, and these distributions are mutually consistent, it then follows that there is a unique probability distribution of the *random sequence* $(x_0, x_1, \cdots)$, the corresponding probability being defined for every Borel set of the space Ω of sequences. Similarly we may consider the doubly infinite sequence $(\cdots, x_{-1}, x_0, x_1, \cdots)$.

Consider further the more general case when **T** is any set of real numbers. Then Ω is the space of all real-valued functions $\omega = x(t)$ defined on the set **T**, and as before the knowledge of the distributions for all finite sets of variables $x(t_1), \cdots, x(t_n)$ permits us to determine a probability distribution in the space Ω of *random functions* $x(t)$, the probability $P(S) = P(x(t) \subset S)$ being always defined for all Borel sets S in Ω.

The generalization of the above considerations to *complex-valued* random sequences and functions is immediate.

6. Various modes of convergence. Consider a sequence $F_1(x), F_2(x)$, of d.f:s, and let the corresponding c.f:s be $\varphi_1(t), \varphi_2(t), \cdots$. In order that $F_n(x)$

converge to a d.f. $F(x)$, in every continuity point of the latter, it is necessary and sufficient[1] that $\varphi_n(t)$ converge for every real t to a limit $\varphi(t)$ which is continuous at $t = 0$. Then $\varphi(t)$ is the c.f. corresponding to the d.f. $F(x)$.

Further, let x and x_1, x_2, $\cdots$ be complex-valued random variables, such that the random sequence $(x, x_1, x_2, \cdots)$ has a well defined distribution. We shall be concerned with various modes of convergence of x_n to x.

A. When $P(\mid x_n - x \mid > \epsilon) \to 0$ as $n \to \infty$, for any $\epsilon > 0$, we shall say that x_n *converges to x in probability*.

B. When $E \mid x_n - x \mid^\gamma \to 0$, as $n \to \infty$, where $\gamma > 0$ is fixed, we shall say that x_n converges to x in the mean of order γ. Unless otherwise stated we shall in the sequel always consider the case $\gamma = 2$, and in this case we shall use the notation

$$\underset{n \to \infty}{\text{l.i.m.}} \, x_n = x.$$

C. When $P(\underset{n \to \infty}{\lim} x_n = x) = 1$, we shall say that x_n converges with probability one, or *converges almost certainly to x*.

With respect to the last definition, we may remark that the set defined by the relation $\lim x_n = x$ is always a Borel set in the space of our random sequence, so that the probability of this relation is well defined. In fact, this probability is given by the expression

$$\lim_{m \to \infty} \lim_{n \to \infty} \lim_{p \to \infty} P\left(\mid x_\nu - x \mid < \frac{1}{m} \quad \text{for} \quad \nu = n, n + 1, \cdots, n + p\right)$$

where the limit process applies to a probability attached to a Borel set in a finite number of dimensions. The case of almost certain convergence is precisely the case when this expression takes the value 1.

Convergence in the mean of any positive order, as well as almost certain convergence, both imply convergence in probability, which may be written symbolically $B \to A$ and $C \to A$. Between B and C, there is no simple relation of this kind. Further, A and B both imply almost certain convergence for any partial sequence x_{n_1}, x_{n_2}, $\cdots$ such that the subscripts n_k increase sufficiently rapidly with k.

II. PROBLEMS CONNECTED WITH THE ADDITION OF INDEPENDENT VARIABLES

7. During the early development of the theory of probability, the majority of problems considered were connected with gambling. The gain of a player in a certain game may be regarded as a random variable, and his total gain in a

---

[1] As I have already stated in a paper published in 1938, there is an error in the statement of this theorem given in my Cambridge Tract [9] *Random Variables and Probability Distributions*. For the truth of the theorem, it is essential that $\varphi_n(t)$ should be supposed to converge to $\varphi(t)$ *for every real t*. However, in the particular case when the limit $\varphi(t)$ is analytic and regular in the vicinity of $t = 0$, it can be proved that it is sufficient to assume convergence in some interval $\mid t \mid < a$.

sequence of repetitions of the game is the sum of a number of independent variables, each of which represents the gain in a single performance of the game. Accordingly a great amount of work was devoted to the study of the probability distributions of such sums. A little later, problems of a similar type appeared in connection with the theory of errors of observation, when the total error was considered as the sum of a certain number of partial errors due to mutually independent causes. At first only particular cases were considered, but gradually general types of problems began to arise, and in the classical work of Laplace several results are given concerning the general problem to study the distribution of a sum

$$z_n = x_1 + \cdots + x_n$$

of independent variables, when the distributions of the x_j are given. This problem may be regarded as the very starting point of a large number of those investigations by which the modern Theory of Probability was created. The efforts to prove certain statements of Laplace, and to extend his results further in various directions, have largely contributed to the introduction of rigorous foundations of the subject, and to the development of the analytical methods. At the same time, more general types of problems have developed from the original problem, and the number and importance of practical applications have been steadily increasing.

8. Composition of distributions. Let x_1 and x_2 be two independent variables, with the d.f.'s F_1 and F_2, and the c.f.'s φ_1 and φ_2, and let the sum $x_1 + x_2$ have the d.f. F and the c.f. φ. Then

$$F(x) = \int_{-\infty}^{\infty} F_1(x - y) \, dF_2(y) = \int_{-\infty}^{\infty} F_2(x - y) \, dF_1(y).$$

We shall say that F is the *composition* of F_1 and F_2, and write this as a symbolical multiplication:

$$F = F_1 * F_2 = F_2 * F_1.$$

To this symbolical multiplication of the d.f:s corresponds a real multiplication of the c.f.'s:

$$\varphi(z) = \varphi_1(z)\varphi_2(z).$$

The operation of composition is both commutative and associative, so that any symbolical product $F = F_1 * F_2 \cdots * F_n$ is uniquely defined and independent of the order of the components. When at least one of the components is continuous (absolutely continuous), the same holds for the composite, and in many cases it is true that the composite is at least as regular as the most regular of the components (Lévy, [58], [63], etc.). However, this general statement does not hold generally, as is shown by an interesting example due to Raikov, [77], where F_1 and F_2 are integral analytic functions, while the composite $F = F_1 * F_2$ is not regular at the origin.

It seems to be an important unsolved problem to find convenient restrictions

ensuring the validity of the above statements of the "smoothing effect" of the operation of composition.

When $F = F_1 * F_2$, we may say that F is "divisible" by each component F_1 and F_2, and it seems natural to try to develop a theory of symbolical factorization for d.f.'s. In this connection, it is important to note that symbolical division is not unique. In fact, Khintchine has shown by an example that it is possible to find the d.f.'s F, F_1, F_2, and F_3 such that

$$F = F_1 * F_2 = F_1 * F_3,$$

while $F_2 \neq F_3$. Another fundamental problem belonging to this order of ideas is to decide whether a given d.f. F is decomposable or not. F is called decomposable, if there is at least one representation of the form $F = F_1 * F_2$, where each component F_ν has more than one point of increase. So far, this problem has only been solved in very special cases, and the general problem still remains open for research. A particular case of some interest would be to know if there exists an absolutely continuous and indecomposable d.f., such that $F(a) = 0$ and $F(b) = 1$ for some finite a and b.

As soon as we restrict ourselves to certain special classes of distributions, it is possible to reach results of a more definite character concerning the factorization problems. Some results of this type will be considered below.

9. Closed families of distributions. The fact that certain families of distributions are closed with respect to the operation of composition has played an important part in many applications. If F_1 and F_2 belong to a family of this character, so does the symbolical product $F = F_1 * F_2$. We first give some simple examples of such families.

The normal distribution. The d.f. F has the form $F = \phi\left(\dfrac{x - m}{\sigma}\right)$, where $\sigma > 0$, and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t^2/2)} \, dt.$$

The c.f. corresponding to F is $e^{miz - \frac{1}{2}\sigma^2 z^2}$, and it follows that for any real m_1, m_2 and any positive σ_1, σ_2 we have

$$\phi\left(\frac{x - m_1}{\sigma_1}\right) * \phi\left(\frac{x - m_2}{\sigma_2}\right) = \phi\left(\frac{x - m}{\sigma}\right),$$

where

$$m = m_1 + m_2, \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

The Poisson distribution. Here the d.f. is $F = F(x; \lambda, m, a)$ where $\lambda > 0$, $a \neq 0$, and F is a step-function with a jump equal to $\dfrac{\lambda^\nu}{\nu!} e^{-\lambda}$ in the point $x = m + \nu a$, where $\nu = 0, 1, \cdots$. The corresponding c.f. is $e^{miz + \lambda(e^{aiz} - 1)}$, and it follows that for any fixed a we have

$$F(x; \lambda_1, m_1, a) * F(x; \lambda_2, m_2, a) = F(x; \lambda_1 + \lambda_2, m_1 + m_2, a).$$

The Pearson Type III distribution. $F = F(x; \alpha, \lambda) = \dfrac{\alpha^\lambda}{\Gamma(\lambda)} \displaystyle\int_0^x t^{\lambda-1} e^{-\alpha t}\, dt,$

$(x > 0)$. The corresponding c.f. is $\left(1 - \dfrac{iz}{\alpha}\right)^{-\lambda}$, and for any fixed $\alpha > 0$ and any

positive λ_1 and λ_2 we have

$$F(x; \alpha, \lambda_1) * F(x; \alpha, \lambda_2) = F(x; \alpha, \lambda_1 + \lambda_2).$$

Stable distributions. We shall say that a closed family is stable, when all its members are of the form $F(ax + b)$, where F is a d.f., while $a > 0$ and b are constants. Obviously the normal family is an example of a stable family. It has been shown by Lévy and Khintchine [49], that a d.f. $F(x)$ generates a stable family when and only when the logarithm of its c.f. is of the form

$$(9.1) \qquad \log \varphi(z) = \beta i z - \gamma \,|z|^\alpha \left(1 + i\delta \frac{z}{|z|} \omega\right),$$

where $\alpha, \beta, \gamma, \delta$ are real constants such that

$$0 < \alpha \leqq 2, \qquad \gamma > 0, \qquad |\delta| \leqq 1,$$

while

$$\omega = \begin{cases} tg \dfrac{\alpha\pi}{2} & \text{for} \quad \alpha \neq 1 \\[2mm] \dfrac{2}{\pi} \log |z| & \text{for} \quad \alpha = 1. \end{cases}$$

For $\alpha = 2$ we obtain the normal family.

A more general and very important closed family is the family I of *infinitely divisible distributions.* A d.f. F belongs to I if to every $n = 1, 2, \cdots$ there exists a d.f. G such that $F = G^{[n]}$, where $G^{[n]}$ denotes the symbolical nth power of G. Obviously the family I is a closed family which contains all the families mentioned above. Lévy [60], [63], has shown that F is infinitely divisible when and only when the logarithm of its c.f. is of the form

$$(9.2) \qquad \begin{aligned} \log \varphi(z) = \beta i z - \gamma z^2 &+ \int_{-\infty}^0 \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dM(u) \\ &+ \int_0^\infty \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dN(u), \end{aligned}$$

where β and $\gamma > 0$ are real constants, while $M(u)$ and $N(u)$ are non-decreasing functions such that

$$M(-\infty) = N(+\infty) = 0,$$

$$\int_{-a}^0 u^2\, dM(u) < \infty \quad \text{and} \quad \int_0^a u^2\, dN(u) < \infty$$

for any finite $a > 0$. When M and N reduce to zero, we obtain the normal family. When $\gamma = 0$ and one of the functions M and N reduces to zero, while

the other is a step-function with a single jump equal to λ at the point $x = a$, we obtain a Poisson family. Generally, it follows from (9.2) that any infinitely divisible distribution may be regarded as a product of a normal distribution and a finite, enumerable or continuous set of Poisson distributions.

The representation of $\log \varphi(z)$ in the form (9.2) is unique. It follows that the problem of finding all possible factorizations of an infinitely divisible d.f. F can be completely solved, as long as we restrict ourselves to factors which are themselves infinitely divisible. In fact, in order that

$$F = F_1 * F_2,$$

where all three d.f.'s belong to I, it is necessary and sufficient that the logarithms of the corresponding c.f.'s should be of the form (9.2), with

$$\beta = \beta_1 + \beta_2, \qquad \gamma = \gamma_1 + \gamma_2,$$
$$M = M_1 + M_2, \qquad N = N_1 + N_2.$$

In the two simple cases of the normal and the Poisson distributions, the decompositions obtained in this way remain the only possible, even if we remove the restriction that the factors should belong to I. Thus in any factorization of a normal distribution, all factors are normal (Cramér, [8]), while in any factorization of a Poisson distribution, all factors belong to the Poisson family (Raikov, [75]). For the type III distribution, and the non-normal stable distributions, however, the corresponding property does not hold.

In some cases, an infinitely divisible distribution may be represented as a product of indecomposable distributions, or as a product of an indecomposable distribution and another infinitely divisible distribution. The results so far obtained in this direction (Lévy, [63], [64], Khintchine, [46], [47]; Raikov, [76]) are all concerned with more or less particular cases, and the general factorization problem for infinitely divisible distributions still remains unsolved. A particular case of some interest would be the case when the functions M and N are both absolutely continuous. There does not seem to have been given any example of this type, where a factor not belonging to I may occur.[2]

Finally we mention a general theorem due to Khintchine, [46], which asserts that an arbitrary d.f. F may be represented in one of the forms

$$F = G, \qquad F = H \text{ or } F = G * H,$$

where G is infinitely divisible, while H is a finite or infinite product of indecomposable factors. This seems to be practically the only result so far known concerning the factorization of a general distribution.

A certain number of the results mentioned above have been generalized to multi-dimensional distributions.

---

[2]While the present paper was being printed, I have proved that such factors do occur, as soon as at least one of the derivatives M' and N' is bounded away from zero in some interval $(-a, 0)$ or $(0, a)$.

10. The Laws of large numbers. In modern terminology, the classical Bernoulli theorem may be expressed in the following way. Let x_1, x_2, $\cdots$ be a sequence of independent variables, such that each x_ν may only assume the values 1 and 0, the corresponding probabilities being p and $q = 1 - p$. Then the arithmetic mean

$$(10.1) \qquad \frac{z_n}{n} = \frac{x_1 + \cdots + x_n}{n}$$

converges in probability to p, as $n \to \infty$.

Both classical and modern authors have laid down much work on the generalization of this simple result in various directions. Generally, we shall say that a sequence of random variables $x_1, x_2, \cdots$ satisfies the *Weak Law of Large Numbers* if there exist two sequences of constants $a_1, a_2, \cdots$ and $b_1, b_2, \cdots$, such that $a_n > 0$, and

$$\frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n}$$

converges in probability to zero.

Let $x_1, x_2 \cdots$ be independent variables, such that x_ν has the d.f. $F_\nu(x)$. It has been shown by Feller [27] that *for any given sequence $a_1, a_2, \cdots$, the conditions*

$$(10.2) \qquad \begin{aligned} &\sum_{\nu=1}^{n} \int_{|x| > a_n} dF_\nu(x) = o(1), \\ &\sum_{\nu=1}^{n} \int_{|x| < a_n} x^2 \, dF_\nu(x) = o(a_n^2), \end{aligned}$$

are sufficient for the validity of the weak law of large numbers, and that the corresponding sequence $b_1, b_2, \cdots$ can be defined by

$$b_n = \sum_{\nu=1}^{n} \int_{|x| < a_n} x \, dF_\nu(x).$$

When there is a constant $c > 0$ such that for all ν

$$(10.3) \qquad F_\nu(+0) > c, \qquad F_\nu(-0) < 1 - c,$$

the conditions are also necessary. This theorem contains as particular cases all previously known results in this direction. A simple NS condition for the existence of at least one sequence $a_1, a_2, \cdots$ such that 10.2 holds does not seem to be known.

When the weak law is satisfied, this means that, for any given $\epsilon > 0$ and *for any fixed large n*, there is a probability very near to 1 that the sum $z_n = x_1 + \cdots + x_n$ will fall between the limits $b_n \pm \epsilon a_n$. The more stringent condition that, with a probability tending to 1 as $n \to \infty$, z_ν will fall between the limits $b_\nu \pm \epsilon a_\nu$ *for all values of $\nu \geqq n$* is equivalent to the condition that $\dfrac{z_n - b_n}{a_n}$ con-

verges *almost certainly* to zero. When this holds, we shall say that the variables x_r satisfy the *Strong Law of Large Numbers*. The most important result so far known in this connection is concerned with the case $a_n = n$, and is expressed by the following theorem (Kolmogoroff, [52], [55]):

When the x_r are independent and (10.3) holds, a sufficient condition for the validity of the strong law with $a_n = n$ consists in the simultaneous convergence of the two series

$$\Sigma \int_{|x| > n} dF_n(x) \qquad and \qquad \Sigma \frac{1}{n^2} \int_{|x| < n} x^2 \, dF_n(x).$$

Some improved conditions of this type have been given by Marcinkicwicz and Zygmund, [73], but the problem of finding a NS condition for the strong law is still unsolved, even in the case $a_n = n$.

Important generalizations of the laws of large numbers to cases when the x_r are not assumed to be independent have been given i.a. by Khintchine [44], Lévy [62], [63] and Loève [67].

11. The central limit theorem and allied theorems. It was already known to De Moivre that, in the case 10.1 of the Bernoulli distribution, the d.f. of the normalized sum

$$\frac{x_1 + \cdots + x_n - np}{\sqrt{npq}}$$

tends, as $n \to \infty$, to the normal d.f. $\phi(x)$. Considerably more general results in this direction were stated by Laplace. After a long series of more or less successful attempts, a rigorous proof of the main statements of Laplace was given in 1901 by Liapouncff, [65]. More general cases were later considered i.a. by Lindeberg [66], Lévy [61], [63], Khintchine [43] and Feller, [25]. The following final form of the *Central Limit Theorem* is due to Feller.

Consider the expression

(11.1) $$u_n = \frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n},$$

where the x_r are independent variables. We shall say that the x_r obey the central limit law, if the sequences $\{a_r\}$ and $\{b_r\}$ can be found such that the d.f. of u_n tends to $\phi(x)$ as $n \to \infty$. In order to avoid unnecessary complications, we shall restrict ourselves to sequences $\{a_r\}$ such that

$$a_r \to + \infty, \qquad \frac{a_{r+1}}{a_r} \to 1,$$

and we shall assume that the conditions (10.3) are satisfied. Then Feller's theorem runs as follows:

The independent variables x_1, x_2, $\cdots$ obey the central limit law if, and only if, there exists a sequence $q_n \to \infty$ such that simultaneously

$$\sum_{\nu=1}^{n} \int_{|x|>q_n} dF_\nu(x) \to 0,$$

(11.2)

$$\frac{1}{q_n^2} \sum_{\nu=1}^{n} \int_{|x|<q_n} x^2 \, dF_\nu(x) \to \infty.$$

When these conditions are satisfied, explicit expressions for the a_n and b_n can be obtained.

Feller's theorem gives a complete solution of the problem. However, we might still try to express in a more direct way the condition that the q_n should exist. We may also ask what happens when the conditions (11.2) are not satisfied. Some particular cases of the latter question will be considered below. However, very few general results are known in this direction.

The central limit theorem has been extended in various directions. Bernstein [3], Lévy [62], [63], Loève [67] and others have considered cases where the x_ν are not assumed to be independent. Important results have been reached but still much remains to be done.

On the other hand, several authors have considered symmetrical functions, other than sums, of n independent random variables. The problem of investigating the asymptotic behaviour of the distributions of such functions, as n tends to infinity, is of great importance in the theory of statistical sampling distributions. It is known (c.f. e.g. Cramér, [15]) that under certain general regularity conditions there exists a normal limiting distribution. However, it is also known that it is possible to give examples of particular functions (such as e.g. the function which is equal to the largest of the n variables), where there exist limiting distributions which are non-normal. The conditions under which this phenomenon may occur seem to deserve further study.

A further problem belonging to the same order of ideas is to find a closer asymptotic representation of the d.f. of the standardized sum z_n than that provided by the normal function $\phi(x)$. Consider e.g. the simple case when the x_ν are independent variables all having the same d.f. $F(x)$ with a finite mean m, a finite variance σ^2, and finite moments up to a certain order $k \geqq 3$. Let $G_n(x)$ be the d.f. of the variable

$$\frac{x_1 + \cdots + x_n - nm}{\sigma \sqrt{n}} .$$

It then follows from a theorem of Cramér [5], [9] that, as soon as the d.f. $F(x)$ contains an absolutely continuous component, there is an asymptotic expansion

(11.3) $$G_n(x) = \phi(x) + \sum_{\nu=1}^{k-3} \frac{p_\nu(x)}{n^{\nu/2}} e^{-x^2/2} + O(n^{-(k-2)/2}),$$

where the constant implied by the O is independent of n and x. Cramér has also given similar expansions in more general cases, and his results have been

further extended by P. L. Hsu [39], who deduces analogous expansions also for other functions than sums. The most general conditions under which expansions of this type exist are still unknown.

It follows from (11.3) that the difference $G_n(x) - \phi(x)$ is, for any fixed x, of the order $n^{-\frac{1}{2}}$ as $n \to \infty$. It is often important to know the asymptotic behaviour of $G_n(x)$ when n and x increase simultaneously, and in that case (11.3) yields only a trivial result. This case has been investigated by Cramér [10], and Feller [29], and the results so far obtained permit important applications to the so called law of the iterated logarithm (cf. below). However, it seems likely that similar results may be obtained in considerably more general cases than those hitherto investigated.

A further interesting type of problems belonging to this order of ideas may be approached in the following way. Consider the variables (11.1) in the particular case when x_1, x_2, $\cdots$ are independent variables all having the same d.f. $F(x)$. When the a_n and b_n can be found such that the d.f. of the normalized sum u_n tends to $\phi(x)$, we shall say that F belongs to the *domain of attraction* of the normal law. Feller's theorem gives a NS condition that this should be so. Now when this condition is not satisfied, it may still occur that the a_n and b_n can be so chosen that the d.f. of u_n tends to a limiting d.f. $\Psi(x)$, which is necessarily different from $\phi(x)$. Then it is easily seen that $\Psi(x)$ must be a stable distribution, with its c.f. defined by (9.1), and it is natural to say that F belongs to the domain of attraction of Ψ. NS and sufficient conditions that this should hold have been given by Doeblin [16], and Gnedenko [34]. When the a_n and b_n cannot be found such that the d.f. of the normal sum u_n converges to a limit, it may still be possible to obtain a limiting d.f. by considering only a partial sequence u_{n_1}, u_{n_2}, $\cdots$. Khintchine [47] has proved the interesting theorem that the totality of limiting d.f.'s that may be obtained in this way coincides with the class of infinitely divisible d.f.'s defined by (9.2). There are also further results in the same direction given by Bawly [2], Khintchine [44], Lévy, [61]-[63], and Gnedenko, [35].

12. The law of the iterated logarithm. Consider a sequence of independent variables x_1, x_2, $\cdots$, such that the mean $Ex_n = 0$ for all n, while the variances $Ex_n^2 = \sigma_n^2$ are finite. Put $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$, and suppose that the variables obey the central limit law with $a_n = s_n$, $b_n = 0$. (In particular this will be the case when all x_n have the same distribution.) For any function $\psi(n)$ tending to infinity with n we then have

$$(12.1) \qquad \lim_{n \to \infty} P(|z_n| > s_n \psi(n)) = 0.$$

On the other hand, if $\psi(n)$ tends to a finite limit > 0, the same probability has a positive limit.

It seems natural to consider the relation within the brackets in (12.1) not only for a single large value of n, but to require the probability that this relation

holds simultaneously for *an infinite number of values of n*. The development of this problem has led to the so called law of the iterated logarithm.

We shall in this respect use the following terminology due to Lévy. A non-decreasing positive function $\psi(n)$ will be said to belong to the *lower class* with respect to the variables x_n if, with a probability equal to one, there are infinitely many n such that

$$|z_n| > s_n \psi(n).$$

On the other hand, $\psi(n)$ will be said to belong to the *upper class* if the probability of the same property is equal to zero.

Every $\psi(n)$ belongs to one of these two classes. This is a special case of the so called *null-or-one law*: if S is a Borel set in the space of the independent random variables $x_1, x_2, \cdots$, such that any two points differing at most in a finite number of coordinates either both belong to S or both belong to the complementary set, then $P(S)$ can only assume the values 0 or 1.

It was proved by Kolmogoroff [51] that, subject to certain restrictions, the function

$$\psi(n) = \sqrt{c \log \log s_n}$$

belongs to the lower class for any $c < 2$, and to the upper class for any $c > 2$, which may be expressed by the relation

(12.1) $$P\left(\limsup_{} \frac{z_n}{s_n \sqrt{2 \log \log s_n}} = 1\right) = 1.$$

More general results were proved by Feller [30], who proved i.a. that, subject to certain restrictions, $\psi(n)$ belongs to the lower or upper class according as

(12.2) $$\Sigma \frac{\sigma_n^2}{s_n^2} \psi(n) e^{-(\psi^2(n)/2)}$$

is divergent or convergent (in certain special cases, this had been previously found by Kolmogoroff and Erdös [24]. Feller also proved a more complicated result, which contains the above as a particular case, and from which it follows that the simple criterion (12.2) no longer holds when the restrictions imposed in its proof are removed.

13. Convergence of series. For any sequence of random variables x_n, the probability

$$P\left(\sum_1^\infty x_n \text{ converges}\right)$$

has a uniquely determined value. When the x_n are independent, it follows from the null-or-one law that this probability is either 0 or 1. By a theorem of Khintchine and Kolmogoroff [48], the value 1 is assumed when and only when the three series

$$\sum_n \int_{|x_n|>1} dF_n, \qquad \sum_n E y_n, \qquad \sum_n \sigma^2 y_n$$

are convergent, where

$$y_n = \begin{array}{ll} x_n & \text{when} \quad |x_n| \leqq 1. \\ 0 & \text{when} \quad |x_n| > 1. \end{array}$$

For the case when the x_n are not assumed to be independent, various results have been given by Lévy [63] and others, but our knowledge of the properties of these series is still not very advanced.

14. Generalizations. In several instances it has been pointed out above that the results concerning sums of independent variables may, to a certain extent, be extended to cases when the variables are not independent. Generally the independence condition has then to be replaced by some condition restricting the degree of dependence. Results of this type were first give by Bernstein [3], and then in more general cases by Lévy [62], [63], and Loève [67]. However, this field has so far only been very incompletely explored.

Similar remarks apply to the generalization of the various theorems quoted above to cases of variables and distributions in more than one dimension.

III. STOCHASTIC PROCESSES

15. The theory of random variables in a finite number of dimensions is able to deal adequately with practically all problems considered in classical probability theory. However, during the early years of the present century, there appeared in the applications various problems, where it proved necessary to consider probability relations bearing on infinite sequences of numbers, or even on functions of a continuous variable.

The mathematical set-up required for the study of such problems involves the introduction of probability distributions in spaces of random sequences or random functions (cf. 5 above). Generally, any process in nature which can be analyzed in terms of probability distributions in spaces of these types will be called a *stochastic process*. It is convenient to apply this name also to the probability distribution used for the study of the process. We shall thus say, e.g., that a certain random function $x(t)$ is attached to the stochastic process which is defined by the probability distribution of $x(t)$. In the majority of applications, the variable t will represent the time, and we shall often use a terminology directly referring to this case. However, there are also other types of problems in the applications (t may e.g. be a spatial variable in an arbitrary number of dimensions), and it is obvious that the purely mathematical problems connected with these classes of probability distributions will have to be considered quite independently of any concrete interpretation of the variable t or the funcion $x(t)$.

A well-known example of this type of problems is afforded by the Brownian movement. Let $x(t)$ be the abscissa at the time t of a small particle immersed in a liquid, and subject to molecular impacts. In every instant, the quantity $x(t)$ receives a random impulse, and the problem arises to study the behaviour of $x(t)$. According as we are content to consider $x(t)$ for a discrete sequence of t-points, say for $t = 0, 1, 2, \cdots$, or we wish to consider all positive values of t,

we shall then have to introduce a probability distribution in the space of the random sequence $x(0)$, $x(1)$, $\cdots$, or in the space of the random function $x(t)$, where $t > 0$. We may then discuss such questions as the distribution of $x(t)$ for a given value of t, the joint and conditional distributions of $x(t)$ for two or more values of t, and, in the case of a continuous variable t, continuity, differentiability and other similar properties of the random function $x(t)$.

Wiener [82], [83] (cf. also Paley and Wiener [74]) was the first to give a rigorous treatment of this process. He proved in 1923 that it is possible to define a probability distribution in a suitably restricted functional space, such that the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(t)$ for any $\Delta t > 0$. With a probability equal to 1, the function $x(t)$ is continuous for all $t > 0$, and for any fixed $t > 0$, the random variable $x(t)$ is normally distributed.

Another example of stochastic processes studied at this stage occurs in the theory of risk of an insurance company. Let $x(t)$ denote the total amount of claims up to the time t in a certain insurance company. As in the case of the Brownian movement, it may seem natural to assume that the increment $\Delta x(t)$ is independent of $x(t)$. On the other hand, $x(t)$ is in this case an essentially discontinuous function, which is never decreasing, and increases only by jumps of varying magnitudes occurring for certain discrete values of t, which are not a priori known. Processes of this type were studied by F. Lundberg [69], [70], H. Cramér [6] and others.

Further examples of particular processes were discussed in connection with various applications, but no general theory of the subject existed until 1931, when Kolmogoroff published a basic paper [53] dealing with the class of stochastic processes which will here be denoted as Markoff processes (Kolmogoroff uses the term "stochastically definite processes"), of which the two examples mentioned above form particular cases. The theory of this class of processes was further developed by Feller [26], [28]. In 1934, Khintchine [42] introduced another important class of processes known as stationary processes. From 1937, the general theory of the subject was subjected to a penetrating analysis in a series of important works by Doob [18]–[22].[3]

16. Probability distributions in functional spaces. We have seen in 5 above how a probability distribution in the space of all functions $x(t)$ may be defined, when t varies in an arbitrary space T. Generally, we shall here content ourselves to consider the cases when T is the set of all real numbers, or the set of all non-negative real numbers. Most results obtained for these cases will be readily generalized to cases when t varies in a Euclidean space of a finite number of dimensions. On the other hand, when T is enumerable, say consisting of the points $t = 0, \pm 1, \pm 2, \cdots$, so that we are concerned with a random sequence $x(0)$, $x(\pm 1)$, $\cdots$, the results for the continuous case will generally hold and assume a simpler form which will not be particularly stated here.

[3] A further interesting paper by Doob has appeared while the present paper was being printed: "Probability in function space", *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 15–30.

The case when T is a space of an infinite number of dimensions does not seem to have been considered so far.

In the present paragraph, it will be convenient to assume the function $x(t)$ to be real-valued, but the generalization to a complex-valued $x(t)$ requires only obvious modifications. In the sequel we shall sometimes consider the real-valued and sometimes the complex-valued case, according as the occasion requires.

Let now X be the space of all real-valued functions $x(t)$ of the real variable t, where $-\infty < t < \infty$. According to 5, a probability measure $P(S)$ is uniquely defined for all Borel sets S in X by means of the family of joint distributions of all finite sequences $x(t_1), \cdots, x(t_n)$. In fact, $P(S)$ can be defined for a more general class of sets than the Borel sets. For any set S in X, we may define an outer P-measure $\overline{P}(S)$ as the lower bound of $P(Z)$ for all sums Z of finite or enumerable sequences of intervals, such that $S \subset Z$. Further, the inner P-measure $\underline{P}(S)$ is defined by the relation $\underline{P}(S) = 1 - \overline{P}(X - S)$. When the outer and inner measures are equal, S is called P-measurable, and $P(S)$ is defined as their common value. Any P-measurable set differs from a Borel set by a set of P-measure zero.

In many cases, this definition will be sufficient for an adequate treatment of the problems that we wish to consider. However, in other cases we encounter certain characteristic difficulties, which make it desirable to consider the possibility of amending the basic definition. Thus it often occurs that we are interested in the probability that the random function $x(t)$ satisfies certain regularity conditions in a non-enumerable set of points t. We may, e.g., wish to consider the probability that $x(t)$ is continuous for all t, that $x(t)$ should be Lebesque-measurable for all t, that $x(t) \leq k$ for all t, etc. Let S denote the set of all functions satisfying a condition of this type. It can then be shown that the inner measure $\underline{P}(S)$ is always equal to zero so that S is never measurable, except in the (usually trivial) case when $P(S) = 0$.

Consequently many interesting probabilities are left undetermined by the general definition of a probability distribution in X given above. The possibility of modifying the definition so as to enable us to study probabilities of this type has been thoroughly investigated by Doob [18]. He considers a subspace X_0 of the general functional space X, where X_0 is chosen so as to contain only, or almost only, "desirable" functions, i.e. functions satisfying such regularity conditions as seem natural with respect to the problem under investigation. We start from a given probability measure $P(S)$ in X, and ask if it is possible to define a probability measure in the restricted space X_0, which corresponds in some natural way to the given distribution in X. Let S_0 be a set in X_0, and suppose that it is possible to find a P-measurable set S in X such that $SX_0 = S_0$. According to Doob, a probability measure P_0 in X_0 is then uniquely defined by the relation

$$P_0(S_0) = P(S)$$

if and only if the condition

$$\overline{P}(X_0) = 1$$

is satisfied.

The problem is thus reduced to finding a subspace X_0 of outer P-measure 1, such that X_0 contains only functions of sufficiently regular behaviour. When this can be done, we can restrict ourselves to consider only functions $x(t)$ belonging to X_0, the probability distribution in this space being defined by the measure P_0. We shall then say that $x(t)$ is a random function, attached to a stochastic process with the restricted space X_0. Doob has obtained a great number of interesting results in this connection, e.g. with respect to the problem of choosing X_0 such that it contains almost only Lebesgue-measurable functions, or such that the probability of the relation $x(t) \leqq k$ has a well-defined value for all k. In particular he has shown that the last problem can be solved for any given P-measure. However, our knowledge of the various possibilities which exist with respect to the choice of X_0 is still very incomplete, and it seems likely that further important results may be reached along this line of research.

An alternative method of introducing probability distributions in functional spaces has been used by Wiener [82], [83], (cf. also Paley and Wiener, [74]). Consider a given probability measure Π in an arbitrary space Ω, defined for all sets Σ of an additive class C. Let $x(t, \omega)$ denote a function (real- or complex-valued, as the case may be) of the arguments t (real) and ω (point in Ω), such that $x(t, \omega)$ for every fixed t becomes a C-measurable function of ω. On the other hand, when ω is fixed, $x(t, \omega) = x(t)$ reduces to a function of the real variable t. Let X_0 denote the set of all functions $x(t)$ corresponding in this way to points of Ω. Further, let $S_0 = SX_0$, where S is a Borel set in X, and let Σ denote the set of all points ω such that $x(t, \omega) \subset S_0$. Then Σ belongs to C, and a probability measure P_0 in the functional space X_0 is uniquely defined by the relation

(16.1) $$P_0(S_0) = \Pi(\Sigma).$$

The relations between the two modes of definition have been discussed by Doob and Ambrose [23] who have shown that they are largely equivalent. However, it seems likely that in particular problems the one or the other procedure may sometimes be the more advantageous, and further investigations on this subject seem desirable.

17. Processes with a finite mean square. Consider a stochastic process defined by a probability measure $P(S)$ in the space X of all complex-valued functions $x(t)$ of the real variable t. For any fixed t_0, the random variable $x(t_0)$ is then a complex-valued function of the variable point $x(t)$ in the space X, i.e. a point Q_{t_0} in the space Ω of all complex-valued functions defined on X. When t_0 varies, the point Q_{t_0} describes a "curve" in Ω, which then corresponds to our stochastic process.

Suppose, in particular, that the mean square

$$E \mid x(t) \mid^2 = \int_X \mid x(t) \mid^2 dP$$

is finite for any fixed value of t. This implies that for fixed t the function $x(t)$ belongs to L_2 over X, relative to the probability measure P. The random variable $x(t)$ may then be regarded as an element of the Hilbert space H of all complex-valued functions f belonging to L_2 over X, the inner product (f, g) of two elements f and g being defined by the relation

$$(f, g) = \int_X f\bar{g} \, dP = E(f\bar{g}).$$

The stochastic process to which $x(t)$ is attached then corresponds to a "curve" in H (Kolmogoroff, [56], [57]), so that the well-known theory of Hilbert space is available for the study of the process. In particular, convergence in the usual metric of Hilbert space is equivalent to convergence in the mean of order 2 for random variables.

Let H_x be the smallest closed linear subspace of H which contains all elements of the form $a_1x(t_1) + \cdots + a_nx(t_n)$. If the *covariance function*

$$r(t, u) = (x(t), x(u)) = E(x(t)\overline{x(u)})$$

is continuous for all real values of t and u, then $x(t) \to x(t_0)$ in the mean, as $t \to t_0$, and we shall say that the process $x(t)$ is *continuous*. For any continuous process, H_x is separable. When $g(t)$ is a continuous non-random function of t, and $x(t)$ is attached to a continuous stochastic process, the Riemann-Darboux sums formally associated with the integral

$$\int_a^b g(t)x(t) \, dt$$

are easily shown to tend to a limit y, which is an element of H_x, i.e. a random variable. By definition, we may identify the integral with this variable y, and this integral will possess the essential properties of the ordinary Riemann integral (Cramér, [12]).

The application of the theory of Hilbert space to stochastic processes seems to open very interesting possibilities. Some applications to particular classes of stochastic processes will be mentioned below. Futher important results belonging to this order of ideas will be given in a work by K. Karhunen [40], which is in course of publication.

18. Relations to ergodic theory. There is a close connection between the theory of stochastic processes and ergodic theory. In ergodic theory, as summarized e.g. in the treatise of E. Hopf [38], we consider an arbitrary space Ω, and a probability measure Π, defined for all sets Σ belonging to the additive

class C. We further consider a one-parameter group of one-one transformations of Ω into itself (a "flow" in Ω) such that the transformation corresponding to the parameter value t takes the point $\omega = \omega_0$ into ω_t, while $(\omega_t)_u = \omega_{t+u}$. Let $f(\omega)$ be a given function, defined throughout Ω, and such that $f(\omega_t)$ is C-measurable for every fixed t. The well-known ergodic theorems due to von Neumann, Birkhoff, Khintchine and others are then concerned with the asymptotic behaviours of mean values, which in the classical cases are of the types

$$\frac{f(\omega_0) + f(\omega_1) + \cdots + f(\omega_{n-1})}{n}$$

or

$$\frac{1}{T} \int_0^T f(\omega_t) \, dt,$$

as n or T tends to infinity. (In the case of the latter expression, it is necessary to introduce some additional condition implying measurability in t.)

Writing $x(t, \omega) = f(\omega_t)$, it is seen that to a given transformation group $\omega \to \omega_t$ and a given function $f(\omega)$, there corresponds a stochastic process in the sense of Wiener's definition (cf. 16). The space X_0 of this process consists of all functions $x(t)$ representable in the form $x(t) = f(\omega_t)$, when $\omega = \omega_0$ varies over Ω. The corresponding probability measure P_0 is defined by (16.1).

Thus any of the above-mentioned ergodic theorems may be expressed as a theorem concerning "temporal" mean values of the types

$$\frac{x(0) + x(1) + \cdots + x(n-1)}{n}$$

or

$$\frac{1}{T} \int_0^T x(t) \, dt.$$

If, according to some reasonable convergence definition, we may assign a limit to either of these expressions, as n or T tends to infinity, this limit will be a random variable, and it is important to find conditions which imply that this variable has a constant value for "almost all" functions $x(t)$, i.e. for all $x(t)$ except at most a set of P_0-measure zero.

In the particular case when $x(0)$, $x(1)$, $\cdots$ are independent variables all having the same distribution, the classical ergodic theorems yield simple cases of the laws of large numbers (cf. 10). The mean ergodic theorem of von Neumann gives the weak law, while the Birkhoff-Khintchine theorem gives the strong law. Some more general results belonging to this order of ideas will be mentioned in the sequel.

It will be seen that the two theories are largely equivalent, and it seems likely that further comparative studies of the methods will be of great value to both sides.

19. Markoff processes. Consider now a stochastic process, defined by a probability measure $P(S)$ in the space X of all real-valued functions $x(t)$ of the

real variable t. For any $t_1 < t_2$, there is a certain conditional probability $P(x(t_2) \subset S \mid x(t_1) = a_1)$ of the relation $x(t_2) \subset S$, relative to the hypothesis that $x(t_1)$ assumes the given value a_1. Suppose now that this conditional probability is independent of any additional hypothesis concerning the behaviour of $x(t)$ for $t < t_1$, so that we have e.g. for any $t_0 < t_1 < t_2$ and for any a_0

$$P(x(t_2) \subset S \mid x(t_1) = a_1) = P(x(t_2) \subset S \mid x(t_1) = a_1, x(t_0) = a_0).$$

In this case the process is called a *Markoff process*.

The general theory of this type of processes, which forms a natural generalization of the classical concept of Markoff chains, has been studied in basic works by Kolmogoroff [53] and Feller [26], [28]. Writing

$$P(x(t) \leqq \xi \mid x(t_0) = a_0) = F(\xi; t, a_0, t_0),$$

where $t_0 < t$, F will be the distribution function of the random variable $x(t)$, relative to the hypothesis $x(t_0) = a_0$. Then F satisfies the Chapman-Kolmogoroff equation

$$(19.1) \qquad F(\xi; t, a_0, t_0) = \int_{-\infty}^{\infty} F(\xi; t, \eta, t_1) \, d_\eta F(\eta; t_1, a_0, t_0),$$

which expresses that, starting from the state $x(t_0) = a_0$, the state $x(t) \leqq \xi$ must be reached by passing through some intermediate state $x(t_1) = \eta$, where $t_0 < t_1 < t$. Subject to certain general conditions, it is possible to show that any solution of this equation satisfies certain integro-differential equations, which in some important cases reduce to partial differential equations of parabolic type, and that the d.f. F is uniquely determined by these equations. However, the general conditions mentioned above are in many cases difficult to apply to particular classes of processes, and it would be important to have further investigations concerning these questions.

Markoff processes (not belonging to the subclass of differential processes, which will be considered in the following paragraph) appear in several important applications, e.g. in the theory of cosmic radiation, in certain genetical problems, in the theory of insurance risk etc. In these cases, we are often concerned with the class of *purely discontinuous* Markoff processes, where the function $x(t)$ only changes its value by jumps. If, in addition, there are only a finite or enumerable set of possible values for $x(t)$, the Chapman-Kolmogoroff equation (19.1) reduces to

$$(19.2) \qquad \pi_{ik}(t_0, t) = \sum_j \pi_{ij}(t_0, t_1)\pi_{jk}(t_1, t),$$

where $\pi_{ik}(t_0, t)$ denotes the "transition probability", i.e. the probability that $x(t)$ will be in the kth state at the time t, when it is known to have been in the ith state at the time t_0. In matrix form, this equation may be written

$$(19.3) \qquad \Pi(t_0, t) = \Pi(t_0, t_1)\Pi(t_1, t),$$

where Π denotes the matrix of the π_{ik}.

When only a sequence of discrete values of t are considered, we have here the classical case of Markoff chains, which has received a detailed treatment in the well-known book by Fréchet [32] (cf. also Doob, [19]). The case when t is a continuous variable has been treated by Feller [28], O. Lundberg [71], Arley [1], and other authors. Some of the most important problems of this branch of the subject are concerned with the existence of a unique system of solutions of (19.2) or (19.3), and with the asymptotic behaviour of the solutions for large values of $t - t_0$. Though important results have been reached, there still remains much to be done here, and the same thing holds a fortiori with respect to the analogous problems for general Markoff processes.

20. Differential processes. A particularly interesting case of a Markoff process arises when, for any $\Delta t > 0$, the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(\tau)$ for $\tau \leq t$. The process is then called a *differential process*. Some of the earliest studied stochastic processes belong to this class, which contains in particular the two examples discussed above in 15. Further cases of such processes arise e.g. in the theory of radioactive disintegration and in telephone technique.

Let us suppose that $x(0)$ is identically equal to zero, and that the process is uniformly continuous in probability in every finite interval $0 \leq t \leq T$, i.e. that for any fixed positive ϵ

$$P(\mid x(t + \Delta t) - x(t) \mid > \epsilon) \to 0$$

as $\Delta t \to 0$, uniformly for $0 \leq t \leq T$ Then it follows from the works of Lévy, [60], [63], Khintchine [47] and Kolmogoroff [54] that, for any $t > 0$, the random variable $x(t)$ has an infinitely divisible distribution, with a characteristic function $\varphi(z; t)$ given by (9.2), where $\beta, \gamma, M(u)$ and $N(u)$ may depend on t.

In the particularly important case when the distribution of the increment $x(t + \Delta t) = x(t)$ does not involve t, but only depends on the length Δt of the interval, we say that the process is *temporally homogeneous*, and in this case we have

$$\log \varphi(z; t) = t \log \varphi(z; 1),$$

so that we obtain the general formula for $\varphi(z; t)$ simply by replacing in (9.2) $\beta, \gamma, M(u)$ and $N(u)$ by $t\beta, t\gamma, tM(u)$ and $tN(u)$ respectively.

When $t \to \infty$, or $t \to 0$, the appropriately normalized distribution of $x(t)$ tends, under certain conditions, to a stable distribution (Cramér [7], Gnedenko [36]). When this limiting distribution is normal, there are sometimes even asymptotic expansions analogous to (11.3). Still, the problem of the asymptotic behaviour of the distribution for large t does not seem to be definitely cleared up.

Khintchine [41] and Gnedenko [37] have given interesting generalizations of the law of the iterated logarithm (cf. 12) to processes of the type considered here.

The continuous process discussed in 15 in connection with the Brownian movement corresponds to the temporally homogeneous case when β, $M(u)$ and $N(u)$ all reduce to zero, so that

$$\varphi(z) = e^{-\gamma t z^2},$$

which shows that the distribution of $x(t)$ is normal, with mean zero and variance $2\gamma t$.

On the other hand, in the applications to the theory of insurance risk, γ is zero, while $M(u)$ and $N(u)$ are connected with the distribution of the various magnitudes of claims. In this type of applications, it is often very important to find the probability that $x(t)$ satisfies an inequality of the form

$$x(t) < a + bt$$

for all values of t. It follows from the discussion in 16 that the definition of a probability of this type is somewhat delicate. The problem, which can be regarded as an extended form of the classical problem of "the gambler's ruin," has been solved in certain particular cases. It leads to integral equations, which in the simplest case are of the Volterra, in other cases of the Wiener-Hopf type (Cramér [6], [13], Segerdahl [79], Täcklind [81]).

21. Orthogonal processes. Consider now the case of a complex-valued $x(t)$, and suppose that $E \mid x(t) \mid^2$ is finite for all t. Without restricting the generality, we may assume that $Ex(t) = 0$ for all t.

Suppose now that instead of requiring, as in the case of a differential process, that the variables $x(\tau)$ and $\Delta x(t)$ should be *independent* when $\tau \leqq t$, we only lay down the less stringent condition that these variables should be *non-correlated*, i.e. that

$$E(x(\tau)\overline{\Delta x(t)}) = 0.$$

We then obtain a process which is no longer necessarily of the Markoff type. The condition implies that, for any two disjoint intervals (t_1, t_2) and (t_3, t_4), we have

$$E[(x(t_2) - x(t_1))(\overline{x(t_4)} - \overline{x(t_3)})] = 0,$$

so that the "chords" corresponding to two disjoint "arcs" of the curve in Hilbert space representing the process are always orthogonal (Kolmogoroff [56], [57]). A process of this type may accordingly be called an *orthogonal process*.

For a process of this type we have, writing $E \mid x(t) \mid^2 = F(t)$, $F(t + \Delta t) - F(t) = E \mid x(t + \Delta t) - x(t) \mid^2$, so that $F(t)$ is a never decreasing function of t. If $F(t)$ is bounded for all t, we shall say that the orthogonal process is bounded. For a bounded orthogonal process, the Stieltjes integral

$$\int_{-\infty}^{\infty} g(t) \, dx(t),$$

where $g(t)$ is bounded and continuous, may be defined as the limit in the mean of sums of the form

$$\sum_{\nu} g(t_{\nu})(x(t_{\nu}) - x(t_{\nu-1})).$$

22. Stationary processes. When we are concerned with a process representing the temporal development of a system governed by laws which are invariant under a translation in time, it seems natural to assume that the joint distribution of any group of variables of the form

$$(22.1) \qquad x(t_1 + \tau), \cdots, x(t_n + \tau)$$

is independent of τ. A process satisfying this condition will be called a *stationary* process. If a stochastic process is defined by means of a "flow" $\omega \rightarrow \omega_t$ in a space Ω (cf. 18), the process will be stationary when and only when the corresponding flow is *measure-preserving*, i.e. if the transformation $\omega \rightarrow \omega_t$ changes any C-measurable set S into a set S_t of the same measure.

Under appropriate conditions with respect to the measurability of $x(t)$, the Birkhoff-Khintchine ergodic theorem holds for a stationary process, i.e. there exists a random variable y such that we have

$$(22.2) \qquad P_0 \left(\lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)\, dt = y \right) = 1,$$

where P_0 is the probability measure in a suitably restricted space in the sense of Doob. Further work seems to be required here, in order to make the situation quite clear, also with regard to metric transitivity.

For a stationary process, any finite moment of the joint distribution of the variables (22.1) is obviously independent of τ. Suppose now that we only require that this invariance under translations in time should hold for moments of the first and second order of the joint distributions, which are assumed to be finite. The wider class of processes obtained in this way may be called *stationary of the second order*. Processes of this type have been studied for the first time by Khintchine [42]. We shall assume that $x(t)$ is complex-valued. Without restricting the generality, we may further assume that $Ex(t) = 0$ for all t. The product moment $E(x(t)\overline{x(u)})$ will then be a function of the difference $t - u$:

$$(22.3) \qquad E(x(t)\overline{x(u)}) = R(t - u).$$

Assuming, in addition, that $R(t)$ is continuous at $t = 0$, it follows that $R(t)$ is continuous for all t, and the process is continuous in the sense of 17. It was shown by Khintchine that a NS condition that a given function $R(t)$ should be associated with a second order stationary and continuous process by means of the relation (22.3) is that we should have

$$(22.4) \qquad R(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

for all t, where the spectral function $F(x)$ is real, never decreasing and bounded. In particular, we have

$$F(+\infty) - F(-\infty) = R(0) = E \mid x(t) \mid^2 = \sigma^2.$$

Khintchine's condition for $R(t)$ was generalized by Cramér to the case of an arbitrary number of processes $x_1(t), \cdots, x_n(t)$, such that the product moments $E(x_i(t)\overline{x_j(u)})$ are functions of the difference $t - u$. The corresponding spectral functions $F_{ij}(x)$ are in general complex-valued and of bounded variation. Further, the expression (Cramér, [12])

$$\sum_{i,j=1}^{n} z_i \bar{z}_j \Delta F_{ij},$$

where $\Delta F_{ij} = F_{ij}(b) - F_{ij}(a)$ is, for any $a < b$, a non-negative Hermite form in the variables z_i. This result is closely connected with a theorem on Hilbert space considered by Kolmogoroff and Julia. It is further shown that, to any given functions $F_{ij}(x)$, $(i, j = 1, \cdots, n)$, satisfying these conditions, we can always find n processes $x_1(t), \cdots, x_n(t)$ such that the joint distribution of any set of variables $x_i(t_j)$ is always *normal*, while the covariance functions $R_{ij}(t - u) = E(x_i(t)\overline{x_j(u)})$ are given by the expression

$$R_{ij}(t) = \int_{-\infty}^{\infty} e^{itx} dF_{ij}(x).$$

For a process $x(t)$ which is continuous and stationary of the second order, with $Ex(t) = 0$ for all t, we have the mean ergodic theorem

$$(22.5) \qquad \underset{T\to\infty}{\text{l.i.m.}} \frac{1}{2T} \int_{-T}^{T} e^{-\lambda it} x(t) \, dt = y$$

for any real λ. The random variable y has the mean 0 and the variance $F(\lambda + 0) - F(\lambda - 0)$, where F is the spectral function appearing in (22.4). If λ is a point of continuity for F, it thus follows that $y = 0$ with a probability equal to 1. On the other hand, if λ is a discontinuity, y has a positive variance. Let $\lambda_1, \lambda_2, \cdots$ be all the discontinuities of $F(x)$, and let $\sigma_1^2, \sigma_2^2, \cdots$ be the corresponding saltuses, while $y_1, y_2, \cdots$ are the limits in the mean obtained from (22.5) for $\lambda = \lambda_1, \lambda_2, \cdots$. Then two different y_j are always orthogonal: $E(y_j \bar{y}_k) = 0$ for $j \neq k$, and we have

$$(22.6) \qquad x(t) = \sum_{\nu} y_\nu e^{\lambda_\nu it} + \xi(t),$$

where $E\xi(t) = 0$ and

$$E \mid \xi(t) \mid^2 = \sigma^2 - \sum_{\nu} \sigma_\nu^2.$$

If $F(x)$ is a step-function, we have $\sigma^2 = \sum_{\nu} \sigma_\nu^2$, and it follows that $\xi(t) = 0$ with a probability equal to 1, so that (22.6) gives a "stochastic Fourier expansion" of $x(t)$ (Slutsky, [80]).

Even when $F(x)$ is arbitrary, we can obtain a spectral representation of $x(t)$ generalizing (22.6). In fact, it can be shown (Cramér, [14]) that $x(t)$ can always be represented by a Fourier-Stieltjes integral

$$(22.7) \qquad x(t) = \int_{-\infty}^{\infty} e^{itu}\, dz(u),$$

where $z(u)$ is a random function attached to a bounded orthogonal process (cf. 21), such that

$$E \mid z(u + \Delta u) - z(u) \mid^2 = F(u + \Delta u) - F(u).$$

Conversely, we have

$$(22.8) \qquad z(u + \Delta u) - z(u) = -\int_{-\infty}^{\infty} \frac{e^{-it(u+\Delta u)} - e^{-itu}}{2\pi it}\, x(t)\, dt,$$

so that there is a one-one correspondence between $x(t)$ and $\Delta z(u)$. The integrals (22.7) and (22.8) are defined as limits in the mean, as shown above in 17 and 21. These results are in close correspondence with generalized harmonic analysis for an arbitrary function, as developed by Wiener [83] and Bochner [4]. The spectral representation of a stochastic process has important applications, some of which will be considered in a forthcoming paper by Karhunen [40]. An extension of the spectral representation to a more general class of processes has been given by Loève [68].

When, in particular, the $x(t)$ process is such that the joint distribution of any group of variables $x(t_1), \cdots, x(t_n)$ is normal, it follows that any increment $\Delta z(u)$ is normally distributed. Since two uncorrelated normally distributed variables are always independent, it follows that in this case the $z(u)$ process is a differential process with normally distributed increments. Important results for this case have recently been given by Doob [22].

The properties of continuity, differentiability etc. for processes of the type here considered are still incompletely known, and further work is required. A further group of important unsolved problems are connected with an interesting decomposition theorem by Wold [84], which holds for processes with a discrete time variable. The generalization of this theorem to the continuous case does not seem to have so far been given in a final form.

REFERENCES

[1] N. ARLEY, "On the theory of stochastic processes and their applications to the theory of cosmic radiation," Thesis, Copenhagen, 1943.

[2] G. M. BAWLY, "Ueber eine Verallgemeinerung der Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Rec. Math. (Mat. Sbornik)*, N. S., Vol. 1 (1936), pp. 917–929.

[3] S. N. BERNSTEIN, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math. Ann.*, Vol. 97 (1927), pp. 1–59.

[4] S. BOCHNER, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Ann.*, Vol. 108 (1933), pp. 378–410.

[5] H. CRAMÉR, "On the composition of elementary errors," *Skand. Aktuarietidskr.*, Vol. 11 (1928), pp. 13–74, 141–180.

[6] ———, "On the mathematical theory of risk." Published by the Insurance Company Skandia, Stockholm, 1930.

[7] ———, "Sur les propriétés asymptotiques d'une classe de variables aléatoires," *C. R. Acad. Sci. Paris*, Vol. 201 (1935), pp. 441–443.

[8] ———, "Ueber eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.*, Vol. 41 (1936), pp. 405–414.

[9] ———, *Random variables and probability distributions*, Cambridge Tracts in Math., Cambridge, 1937.

[10] ———, "Sur un nouveau théorème—limite de la théorie des probabilités," *Actualités Scientifiques*, Paris, No. 736 (1938), pp. 5–23.

[11] ———, "On the representation of a function by certain Fourier integrals," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 191 201.

[12] ———, "On the theory of stationary random processes," *Ann. of Math.*, Vol. 41 (1940), pp. 215–230.

[13] ———, "Deux conférences sur la theorie des probabilités," *Skand. Aktuarietidskr.*, 1941, pp. 34-69.

[14] - ———, "On harmonic analysis in certain functional spaces," *Ark. Mat. Astr. Fys.*, Vol. 28B (1942), pp. 1 7.

[15] ———, *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, 1946.

[16] W. DOEBLIN, "Premiers éléments d'une étude systématique de l'ensemble de puissances d'une loi de probabilités," *C. R. Acad. Sci. Paris*, Vol. 206 (1938), pp. 306-308.

[17] ———, "Sur un théorème du calcul des probabilités," *C. R. Acad. Sci. Paris*, Vol. 209 (1939), pp. 742 743

[18] J. L. DOOB, "Stochastic processes depending on a continuous parameter," *Trans. Amer. Math. Soc.*, Vol. 42 (1937), pp. 107 140.

[19] ———, "Stochastic processes with an integral-valued parameter," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 87-150.

[20] ———, "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp. 455–486.

[21] ———, "The law of large numbers for continuous stochastic processes," *Duke Math. Jour.*, Vol. 6 (1940), pp. 290–306.

[22] ———, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.

[23] J. L. DOOB AND M. AMBROSE, "On the two formulations of the theory of stochastic processes depending upon a continuous parameter," *Ann. of Math.*, Vol. 41 (1940), pp. 737–745.

[24] P. ERDÖS, "On the law of the iterated logarithm," *Ann. of Math.*, Vol. 43 (1942), pp. 419–436.

[25] W. FELLER, "Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol. 40 (1935), pp. 521 559.

[26] ———, "Zur Theorie der stochastischen Processe (Existenz- und Eindeutigkeitssatze)," *Math. Ann.*, Vol. 113 (1936), pp. 113–160.

[27] ———, "Ueber das Gesetz der Grossen Zahlen," *Acta. Univ. Szeged.*, Vol. 8 (1937), pp. 191–201.

[28] ———, "On the integro-differential equations of purely discontinuous Markoff processes," *Trans. Amer. Math. Soc.*, Vol. 48 (1940), pp. 488–575.

[29] ———, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.

[30] ———, "The general form of the so-called law of the iterated logarithm," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 373–402.

[31] ———, "The fundamental limit theorems in probability," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 800–832.

[32] M. Fréchet, *Recherches théoriques modernes sur la théorie des probabilités*, Vol. 2, Paris, 1937.

[33] B. V. Gnedenko, "Sur les fonctions caractéristiques," *Bull. Math. Univ. Moscou*, Vol. 1, (1937), pp. 16-17.

[34] ———, "On the theory of the domains of attraction of stable laws," *Učenye Zpiski, Moskovskoga gosudaistvenogo Univaziteta*, Vol. 30 (1939), pp. 61-81.

[35] ———, "On the theory of limit theorems for sums of independent random variables," *Bull. Acad. Sci. URSS*, Vol. 30 (1939), pp. 181-232, 643-647.

[36] ———, "On locally stable probability distributions," *C. R. Acad. Sci. URSS*, Vol. 35 (1942), pp. 263-266.

[37] ———, "Investigation on the growth of homogeneous random processes," *C. R. Acad. Sci. URSS*, Vol. 36 (1942), pp. 3-41.

[38] E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, Vol. 5, No. 2, Berlin, 1937.

[39] P. L. Hsu, "The approximate distribution of the mean and of the variance of independent variates," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 1-29.

[40] K. Karhunen, Paper on stochastic processes, to appear in the *Acta Soc. Sci. Fennicae*.

[41] A. Khintchine. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 4, Berlin, 1933.

[42] ———, "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1934), pp. 604-615.

[43] ———, "Sul dominio di attrazione della legge di Gauss," *Giorn. Ist. Ital. Attuari*, Vol. 6 (1935), pp. 378-393.

[44] ———, "Su una legge dei grandi numeri generalizzata," *Giorn. Ist. Ital. Attuari*, Vol. 7 (1936), pp. 365-377.

[45] ———, "Zur Kennzeichnung der characteristischen Funktionen," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 1-31

[46] ———, "Contribution à l'arithmetique des lois de distribution," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 6-17.

[47] ———, "Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze," *Rec. Math. N. S.*, Vol. 2 (1937), pp. 79-117.

[48] A. Khintchine and A. Kolmogoroff, "Ueber Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Rec. Math.*, Vol. 32 (1925), pp. 668-677.

[49] A. Khintchine and P. Lévy, "Sur les lois stables," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 374-376.

[50] A. Kolmogoroff, "Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen," *Math. Ann.*, Vol. 99 (1928) and Vol. 102 (1929), pp. 484-489.

[51] ———, "Ueber das Gesetz des iterierten Logarithmus," *Math. Ann.*, Vol. 101 (1929), pp. 126-135.

[52] ———, "Sur la loi forte des grands nombres," *C. R. Acad. Sci. Paris*, Vol. 191 (1930), pp. 910-911.

[53] ———, "Ueber die analytischen Methoden der Wahrscheinlichkeitsrechnung," *Math. Ann.*, Vol. 104 (1931), pp. 415-458.

[54] ———, "Sulla forma generale di un processo stocastico omogeneo," *Atti. Acad. naz. Lincei, Rend.*, Vol. 15 (1932), pp. 805-828, 866-869.

[55] ———, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 3, Berlin, 1933.

[56] ———, "Wiener Spiralen und einige andere interessante Kurven im Hilbertschen Raum," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 115-118.

[57] ———, "Kurven im Hilbertschen Raum, die gegenüber einer ein parametrigen Gruppe von Bewegungen invariant sind," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 6-9.

[58] P. Lévy, *Calcul des probabilités*, Paris, 1925.

[59] ———, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Math.*, Vol. 3 (1931), pp. 117-155.

[60] ——————, "Sur les intégrales dont les éléments sont des variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa*, *(2)*, Vol. 3 (1934), pp. 337-366.

[61] ——————, "Propriétés asymptotiques des sommes de variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa (2)*, Vol. 3 (1934), pp. 347-402.

[62] ——————, "La loi forte des grands nombres pour les variables aléatoires enchainées," *Jour. Math. Pures Appl.*, Vol. 15 (1936), pp. 11-24.

[63] ——————, *Théorie de l'addition des variables aléatoires*, Paris, 1937.

[64] ——————, "L'arithmetique des lois de probabilités," *Jour. Math. Pures Appl.*, Vol. 17 (1938), pp. 17-39.

[65] A. M. LIAPOUNOFF, "Nouvelle forme du théorème sur la limite de la probabilité," *Mémoires Acad. Saint-Petersbourg s. 8*, Vol. 12 (1901).

[66] J. W. LINDEBERG, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Zeft.*, Vol. 15 (1922), pp. 211-225.

[67] W. LOÉVE, "Étude asymptotique des sommes de variables aléatoires liées," *Jour. Math. Pures Appl.*, Vol. (9) 24 (1945), pp. 249-318.

[68] ——————, "Analyse harmonique générale d'une fonction aléatoire," *C. R. Acad. Sci. Paris*, Vol. 220 (1945), pp. 380-382.

[69] F. LUNDBERG, "Zur Theorie der Rückversicherung," *Verhandlungen Kongr. f. Versicherungsmathematik*, Wien, 1909.

[70] ——————, *Försäkringsteknisk riskutjämning*, Stockholm, 1927.

[71] O. LUNDBERG, "On random processes and their application to sickness and accident statistics," Thesis, Stockholm, 1940.

[72] J. MARCINKIEWICZ, "Sur une propriété de la loi de Gauss," *Math. Zeit.*, Vol. 4 (1938), pp. 612-618.

[73] J. MARCINKIEWICZ AND A. ZYGMUND, "Sur les fonctions indépendantes," *Fund. Math.*, Vol. 29 (1937), pp. 60-90.

[74] R. E. A. C. PALEY AND N. WIENER, *Fourier Transforms in the Complex Domain*, Amer. Math. Soc. Colloquium Publ., Vol. 19, New York, 1934.

[75] D. RAIKOV, "On the composition of Poisson laws," *C. R. Acad. Sci. URSS*, Vol. 14 (1937), pp. 9-11.

[76] ——————, "On the decomposition of Gauss and Poisson laws," *Bull. Acad. Sci. URSS, Ser. Math.*, 1938, pp. 91-120.

[77] ——————, "On the composition of analytic distribution functions," *C. R. Acad. Sci. URSS*, Vol. 23 (1939), pp. 511-514.

[78] I. J. SCHOENBERG, "Metric spaces and positive definite functions," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 522-536.

[79] C. O. SEGERDAHL, "On homogeneous random processes and collective risk theory," Thesis, Stockholm, 1939.

[80] E. SLUTSKY, "Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationaires en composantes," *Actualités Scientifiques*, No. 738, Paris, 1938, pp. 33-55.

[81] S. TÄCKLIND, "Sur le risque de ruine dans des jeux inéquitables," *Sknad. Aktuarietidskr.*, 1942, pp. 1-42.

[82] N. WIENER, "Differential space," *Jour. Math. Phys. M. I. T.*, Vol. 2 (1923), pp. 131-172.

[83] ——————, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1930), pp. 117-258.

[84] H. WOLD, "A study in the analysis of stationary time series," Thesis, Stockholm, 1938.

THE ESTIMATION OF DISPERSION FROM DIFFERENCES[1]

By Anthony P. Morse[2] and Frank E. Grubbs

Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

Summary. The estimation of variance by use of successive differences of higher order is discussed in this paper. Heretofore, attention has been focused, in published works, on estimates of variance obtained by employing the sum of squares of deviations from the mean and also by using mean square successive differences of the first order [1], [2], [3], [9]. A concise description of the method employing differences of any order with appropriate formulae for the precision of estimates so obtained and also a practical example on the use of the technique are given in section 11. Fundamental contributions to the estimation of variance from higher order differences, a study of the efficiency of the technique and proper orientation of the subject matter in the field of mathematical statistics are given in sections 2–10 of the paper.

1. Introduction. It frequently happens that successive observations, made at regular intervals of time, are subject to the same standard error while the means of the populations from which they are drawn display some kind of trend. The type of trend we speak of is brought about because of the manner in which we have to take measurements or because of variations in the measuring technique itself; or, again, the trend may be characteristic of the thing we are measuring. In any event, we may desire to eliminate the trend in order to study residual effects. As an example, it is desirable in the field of ballistics to evaluate the dispersion of machine guns firing from a moving airplane.

It may also happen that it is either inexpedient or impossible to estimate the standard error of the observations by the method of least squares, for in a large number of cases the type of trend is unknown. In this event a method employing differences of an appropriate order may prove valuable. The method consists merely of arranging the data in a vertical column in the order in which the observations were taken and then forming difference columns in the usual way of order 1, 2, up to say 5 or some other number depending on the peculiarities of the problem at hand and the number of the original observations. Next, sum the squares of the numbers in each column and divide the sum of squares of the pth order differences by $(n - p) \binom{2p}{p}$. When $n \geq 2$ and $p \geq 1$, the numbers thus arrived at are all unbiased estimates of the population variance σ^2 for the case where all the observations have the same expected value. In section 11 at the

end of the paper will be found a summary of this method, formulas by which the precision of the estimate of the variance σ^2 may be determined, and an example displaying the stability of this estimate with respect to p.

If a strong trend is present then the method of first differences will obviously yield an estimate of variance which is fictitiously large and the temptation to pass to higher order differences may quite reasonably be yielded to. As a matter of fact, unbiased estimates may be hoped for from pth order differences whenever there is good reason to suppose that the pth derivative of the trend function is small most of the time. However, even in the case of a sinusoidal trend where all derivatives have the same magnitude one may obtain good results frcm higher differences provided there are at least seven observations in each interval of length one period (see section 5 and Table II below). In connection with trends such as the sinusoidal type, the hopelessness of getting, say, even a fifth degree polynomial to fit over an interval of, say 20 periods is rather evident. It is for the above reasons that estimation of variance from higher order differences deserves consideration.

2. Historical comment. A brief historical development of the interest in successive differences as a means for estimating dispersion is given in [3]. This paper discusses the statistic

$$s^1 = \sqrt{\frac{\sum_{i=1}^{n/2} (x_{1i} - x_{2i-1})^2}{n}}$$

suggested by "Student" [W. S. Gossett] and E. S. Pearson and points out the relevant work of Jordan, Helmert, Vallier, Cranz, and Becker. It seems that Jordan devised methods based on sums of powers of the differences, whereas Helmert gave more careful consideration to the case of the first power, i.e. the sum of absolute differences. Reference [3] points out, however, that in these two cases all the $n(n-1)/2$ differences that can be established frcm a sample of n observations were included in the estimates of dispersion recommended by Jordan and Helmert, so that the estimate was of no value in reducing the effect of a trend. Continuing the remarks of [3], we learn that in ballistics Vallier appears to have been the first to estimate dispersion from successive differences and that Cranz and Becker commended the mean successive difference

$$E_d = \frac{\sum_{i=1}^{n-1} |x_{i+1} - x_i|}{n-1}$$

in estimating dispersion in range of guns since they were aware of variable external effects (such as tail winds) on a projectile. In this country, Bennett [1] appears to have suggested the use of successive differences independently of European ballisticians. In this connection, Bennett suggested that the probable

TABLE I

The Efficiency, $W(n, p)$, of $\delta^2_{n,p}$ As An Estimate of σ^2

p \ n	1	2	3	4	5	6	7	8	9	10
2	1.00000									
3	.80000	.50000								
4	.75000	.46154	.33333							
5	.72727	.46552	.32000	.25000						
6	.71429	.47213	.33149	.24427	.20000					
7	.70588	.47771	.34453	.25510	.19672	.16667				
8	.70000	.48214	.35537	.26871	.20633	.16471	.14286			
9	.69565	.48568	.36408	.28071	.21888	.17274	.14159	.12500		
10	.69231	.48855	.37113	.29071	.23058	.18385	.14830	.12414	.11111	
11	.68966	.49091	.37691	.29904	.24070	.19476	.15802	.12978	.11050	.10000
12	.68750	.49288	.38173	.30602	.24934	.20450	.16798	.13827	.11529	.09955
13	.68571	.49455	.38580	.31194	.25672	.21300	.17714	.14729	12271	.10366
14	.68421	.49598	.38928	.31701	.26308	.22039	.18530	.15581	.13086	.11018
15	.68293	.49722	.39228	.32139	.26859	.22684	.19250	.16353	.13874	.11754
16	.68182	.49831	.39490	.32522	.27342	.23251	.19887	.17045	.14601	.12481
17	.68085	.49926	.39721	.32859	.27767	.23752	.20452	.17664	.15260	.13162
18	.68000	.50011	.39925	.33158	.28145	.24197	.20956	18218	.15855	.13787
19	.67925	.50087	.40107	.33424	.28482	.24595	.21407	.18715	.16393	.14356
20	.67857	.50155	.40271	.33663	.28784	.24953	.21813	.19164	.16879	.14875
21	.67797	.50216	.40419	.33880	.29058	.25276	.22181	.19571	.17321	.15347
22	.67742	.50272	.40553	.34075	.29306	.25569	.22515	.19941	.17723	.15778
23	.67692	.50323	.40675	.34254	.29532	.25837	.22819	.20279	.18091	.16173
24	.67647	.50370	.40787	.34417	.29739	.26082	.23098	.20588	.18428	.16535
25	.67606	.50413	.40889	.34567	.29929	.26307	.23354	.20873	.18738	.16869
26	.67568	.50452	.40984	.34706	.30104	.26514	.23590	.21135	.19024	.17177
27	.67533	.50489	.41071	.34833	.30266	.26705	.23809	.21378	.19289	.17463
28	.67500	.50523	.41152	.34951	.30416	.26884	.24012	.21603	.19535	.17728
29	.67470	.50555	.41228	.35062	.30555	.27049	.24200	21812	.19764	.17975
30	.67442	.50585	.41298	.35165	.30686	.27203	.24375	.22007	.19978	.18205
31	.67416	.50612	.41363	.35260	.30807	.27347	.24539	.22190	.20177	.18420
32	.67391	.50638	.41425	.35350	.30921	.27482	.24693	.22361	.20364	.18622
33	.67368	.50662	.41482	.35434	.31027	.27608	.24837	.22521	.20539	.18811
34	.67347	.50685	.41536	.35513	.31128	.27727	.24973	.22672	.20704	.18989
35	.67327	.50707	.41587	.35588	.31222	.27839	.25101	.22814	.20859	.19157
36	.67308	.50727	.41635	.35658	.31312	.27945	.25221	.22949	.21006	.19315
37	.67290	.50746	.41671	.35724	.31396	.28045	.25335	.23075	.21145	.19465
38	.67273	.50764	.41724	.35787	.31476	.28140	.25443	.23195	.21276	.19606
39	.67257	.50781	.41765	.35847	.31551	.28229	.25545	.23309	.21401	.19741
40	.67241	.50797	.41804	.35904	.31623	.28314	.25642	.23417	.21519	.19868

TABLE I—*Continued*

p \ n	1	2	3	4	5	6	7	8	9	10
42	.67213	.50828	.41875	.36009	.31756	.28472	.25822	.23617	.21738	.20105
44	.67188	.50855	.41941	.36104	.31877	.28615	.25986	23799	.21937	.20320
46	.67164	.50880	.42000	.36191	.31987	.28745	.26135	.23965	.22118	.20516
48	.67143	.50903	.42055	.36271	.32088	.28865	.26271	.24117	.22284	.20695
50	.67123	.50925	.42105	.36343	.32180	.28975	.26397	.24256	.22437	.20860
52	.67105	.50944	.42151	.36411	.32266	.29076	.26512	.24385	.22578	.21012
54	.67089	.50962	.42193	.36473	.32345	.29170	.26619	.24504	.22708	.21153
56	.67073	.50979	.42233	.36531	.32418	.29257	.26718	.24614	.22829	.21284
58	.67059	.50995	.42270	.36585	.32487	.29338	.26811	.24717	.22941	.21405
62	.67033	.51022	.42337	.36682	.32609	.29484	.26977	.24903	.23144	.21624
66	.67010	.51048	.42395	.36767	.32718	.29612	.27123	.25065	.23322	.21817
70	.66990	.51069	.42447	.36843	.32813	.29725	.27252	.25209	.23479	.21987
74	.66972	.51089	.42492	.36910	.32898	.29826	.27368	.25237	.23619	.22138
78	.66957	.51107	.42534	.36970	.32975	.29917	.27471	.25452	.23745	.22274
82	.66942	.51122	.42571	.37024	.33043	29998	.27564	.25556	.23859	.22397
90	.66917	.51150	42636	.37118	.33162	.30139	.27725	.25735	.24055	.22609
98	.66897	.51172	.42689	.37197	.33262	.30257	.27860	.25885	.24219	.22786
106	.66879	.51192	.42735	.37263	.33346	.30357	.27974	.26012	.24358	.22936
114	.66864	.51208	.42774	.37321	.33418	.30443	.28071	.26121	.24477	.23065
122	.66851	.51223	.42808	.37370	33482	.30518	28156	.26216	.24581	.23177
138	.66829	.51247	.42864	.37452	.33585	.30641	.28297	.26372	.24752	.23362
154	.66812	.51266	.42909	.37517	.33667	.30738	.28408	.26496	.24887	.23508
170	.66798	.51281	.42944	.37570	.33734	.30817	28498	.26596	.24997	.23627
202	66777	.51304	43000	.37649	.33836	.30937	.28635	.26749	.25164	.23808
234	.66762	.51322	.43040	.37708	.33909	.31025	.28735	.26860	.25285	.23939
266	.66751	.51335	.43070	.37752	.33965	.31091	.28810	.26944	.25377	.24038
330	.66734	.51353	.43112	.37814	.34044	.31185	.28917	.27063	.25508	.24179
394	.66723	.51365	.43141	.37856	.34097	.31248	.28990	.27143	.25596	.24274
522	.66709	.51381	.43178	.37910	.34164	.31327	.29081	27244	.25707	.24394
778	.66695	.51396	.43215	.37963	.34233	.31408	.29173	.27347	.25819	.24516
1290	.66684	.51409	.43245	.38007	.34288	.31474	.29248	.27430	.25910	.24613
2314	.66676	.51418	.43264	.38036	.34325	.31518	.29298	.27486	.25971	.24680
∞	.66667	.51429	.43290	.38073	.34372	.31573	.29361	.27556	.26048	.24763

error should be estimated from the root mean square successive differences as follows:

$$P.E. = .6745 \sqrt{\frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{2(n-1)}}.$$

In 1940, J. von Neumann and R. H. Kent in [2] investigated further the estimation of probable error from mean square successive differences (sums of squares of first differences). J. von Neumann, R. H. Kent, H R. Bellinson, and B. I. Hart [3] considered the distribution of

$$\delta^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

in a paper which appeared in June 1941. J. D. Williams [4] obtained the moments of $\eta = \dfrac{\delta^2}{s^2}$, where

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

and indicated that the rth moment of η is equal to the rth moment of δ^2 divided by the rth moment of s^2. The distribution of the ratio of the mean square successive difference to the variance has been published by J. von Neumann [5], [6] and B. I. Hart tabulated the probability integral and obtained percentage points for this statistic ([7], [8]). Indeed, it should be remarked that the statistical theory of successive differences is allied with the problem of serial correlation [9]. Finally, the use of squared differences of higher order than the first for estimating variance appears to have been suggested by A. A. Bennett. Quite independently, a treatment of the subject was given by Morse [10] in connection with problems on exterior ballistics. Various results on successive-difference estimation including significance tests have been given by Tintner [13]. One of Tintner's tests involves the use of selected sets of differences.

3. Definitions and notations. Suppose the observations x_1, x_2, x_3, $\cdots x_n$ are made at times $a = t_1 < t_2 < t_3 < \cdots < t_n = b$ and the t_i are uniformly spaced without error. Let $f(t_i)$ be the true trend so that $\eta_i = f(t_i)$ is the mean of the population from which x_i is drawn and $\epsilon_i = x_i - \eta_i$ is a random error. Further, let p be a non-negative integer less than n and denote to the ith backward difference of order p of x by $\Delta^p x_i$, i.e.

$$\Delta^p x_i = \Delta^{p-1} x_i - \Delta^{p-1} x_{i-1} = \sum_{r=0}^{p} (-1)^r \binom{p}{r} x_{i-r},$$

$$\text{where} \quad \binom{m}{n} = \frac{m!}{n!(m-n)!}; \quad \text{and } i \geq p + 1.$$

We define the following:

(1)
$$\delta^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \epsilon_i)^2;$$

(2)
$$d^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p x_i)^2;$$

(3)
$$\nu^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \eta_i)^2;$$

(4)
$$k_{n,p} = \frac{2}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \eta_i)(\Delta^p \epsilon_i).$$

By $E(u)$ we will mean the expected value of u, whereas the variance of u will be denoted by

$$\mathrm{Var}\ (u) = E\{u - E(u)\}^2.$$

Basically, we shall assume that the ϵ_i are sufficiently Gaussian and independent that

$$E(\epsilon_i) = E(\epsilon_i^3) = 0, \qquad E(\epsilon_i^2) = \sigma^2,$$

$$\mu_4 = E(\epsilon_i^4) = 3\sigma^4,$$

$$E(\epsilon_i^\alpha \epsilon_j^\beta) = E(\epsilon_i^\alpha)E(\epsilon_j^\beta),$$

whenever i, j, α and β are positive integers for which

$$i \neq j, \qquad 1 \leq i \leq n, \qquad 1 \leq j \leq n.$$

4. Expected values. We will now determine the mean or expected values of $\delta^2_{n,p}$ and $d^2_{n,p}$.

$$E(\delta^2_{n,p}) = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} E\left\{\sum_{r=0}^{p} (-1)^r \binom{p}{r} \epsilon_{i-r}\right\}^2,$$

$$E(\delta^2_{n,p}) = \frac{1}{\binom{2p}{p}} \sum_{r=0}^{p} \binom{p}{r}^2 \sigma^2.$$

or

(5)
$$E(\delta^2_{n,p}) = \sigma^2 .$$

(see Lemma 1.3 of section 6 below),

Continuing, we have

$$E(d_{n,p}^2) = \frac{1}{\binom{2p}{p}(n-p)} E\left\{ \sum_{i=p+1}^{n} (\Delta^p \epsilon_i + \Delta^p \eta_i)^2 \right\},$$

$$E(d_{n,p}^2) = \frac{1}{\binom{2p}{p}(n-p)} \left\{ (n-p)\binom{2p}{p}\sigma^2 + \sum_{i=p+1}^{n} (\Delta^p \eta_i)^2 \right\},$$

or

(6) $$E(d_{n,p}^2) = \sigma^2 + \nu_{n,p}^2.$$

Consequently, we observe, $d_{n,p}^2$ is on the average larger than σ^2 by the quantity $\nu_{n,p}^2$. In a particular problem, therefore, we are faced with the situation of choosing that combination of n and p which (i) regulates the size of $\nu_{n,p}^2$ and (ii) gives the desired precision of our estimate of variance.

5. The magnitude of $\nu_{n,p}^2$. In order to study the size of $\nu_{n,p}^2$, we will derive for this quantity an upper bound which will indicate the applicability of the method of differences to non-polynomial as well as polynomial trends. Now,

$$\Delta^p \eta_i = \Delta^p f(t_i) = \int_{t_{i-1}}^{t_i} \int_0^h \cdots \int_0^h f^{(p)}(y_1 - y_2 - \cdots - y_p)\, dy_p\, dy_{p-1} \cdots dy_1,$$

where $t_r - t_{r-1} = h$, by straightforward integration. It will be convenient to change the order of integration; thus

$$\Delta^p f(t_i) = \int_0^h \cdots \int_0^h \int_{t_{i-1}}^{t_i} f^{(p)}(y_1 - y_2 - \cdots - y_p)\, dy_1\, dy_p \cdots dy_2.$$

Since, from Schwarz's inequality it is clear that

$$\left\{ \int_\alpha^\beta g(s)\, ds \right\}^2 \le (\beta - \alpha) \int_\alpha^\beta \{g(s)\}^2\, ds$$

whenever α and β are real numbers and g is integrable, we have

$$\{\Delta^p \eta_i\}^2 \le h^p \int_0^h \cdots \int_0^h \int_{t_{i-1}}^{t_i} \{f^{(p)}(y_1 - y_2 - \cdots - y_p)\}^2\, dy_1\, dy_p \cdots dy_2.$$

Also,

$$\sum_{i=p+1}^{n} \{\Delta^p \eta_i\}^2 \le h^p \int_0^h \cdots \int_0^h \int_{t_p}^{t_n} \{f^{(p)}(y_1 - y_2 - \cdots - y_p)\}^2\, dy_1\, dy_p \cdots dy_2.$$

But for $0 \le r \le (p-1)h = t_p - a$ we have

$$\int_{t_p}^{t_n} \{f^{(p)}(y_1 - r)\}^2\, dy_1 = \int_{t_p-r}^{t_n-r} \{f^{(p)}(s)\}^2\, ds \le \int_a^b \{f^{(p)}(s)\}^2\, ds.$$

Consequently

$$\sum_{i=p+1}^{n} (\Delta^p \eta_i)^2 \le h^p \int_0^h \cdots \int_0^h \int_a^b \{f^{(p)}(s)\}^2 \, ds \, dy_p \cdots dy_2 = h^{2p-1} \int_a^b \{f^{(p)}(s)\}^2 \, ds.$$

Since $h = \dfrac{b-a}{n-1}$, we have finally

(7) $$\nu_{n,p}^2 \le \frac{1}{\binom{2p}{p}} \left(\frac{b-a}{n-p}\right)\left(\frac{b-a}{n-1}\right)^{2p-1} \int_a^b \frac{\{f^{(p)}(s)\}^2 \, ds}{b-a},$$

which is an upper bound for $\nu_{n,p}^2$ in terms of the average value of the square of the pth derivative of the trend function f.

If the trend function f is of the polynomial form,

$$f(t) = \sum_{r=0}^{p} a_r t^r$$

then the effect of the trend can be eliminated from our observations by estimating dispersion from $(p+1)$st differences. However, if it is *known* that the trend is of polynomial form, then an estimate of dispersion based on least squares would, of course, be better. In fact, it will be shown later that the precision of $\delta_{n,p}^2$ decreases markedly as p increases. The use of $d_{n,p}^2$ as an estimate of σ^2 is primarily of value when the type of trend is unknown; however, even when the type of trend is known the computational simplicity of $d_{n,p}^2$ may offset to some extent its lack of optimum precision.

Let us reflect on the magnitude of $\nu_{n,p}^2$ over a single period of a sinusoidal trend, say $f(t) = \sin t$. In (7) we set $a = 0$, $b = 2\pi$ and secure

$$\nu_{n,p}^2 \le \frac{\pi}{\binom{2p}{p}(n-p)} \left(\frac{2\pi}{n-1}\right)^{2p-1}.$$

Taking n to be the number of observations for a complete period, a tabulation of the upper bound for $\nu_{n,p}^2$ for this case is given in Table II. Thus, when there are about seven or more observations in each interval of length one period, estimation of dispersion from higher order differences may prove of considerable value even for this rather extreme type of trend.

6. Some combinatorial relations. Although we will ultimately establish expressions for the variances of $\delta_{n,p}^2$ and $d_{n,p}^2$, it appears desirable to give first a number of combinatorial relations which present themselves in the computation of moments. The relations are easily checked and most of them are possibly well known. Nevertheless, it will be convenient to record them for reference and in some instances to give proofs. In what follows it will be understood that $\binom{p}{q} = 0$ whenever p and q are not such integers that $0 \le q \le p$.

TABLE II

ⁿ ⧸ ⱼ	5	6	7	8	9	10
1	.617	.395	.274	.201	.154	.110
2	.676	.260	.120	.063	.036	.016
3	.751	.164	.049	.018	.008	.002
4	.106	.111	.021	.005	.002	.0003
5	—	.098	.009	.002	.0004	.0000

LEMMA 1.1. $q\begin{pmatrix} p \\ q \end{pmatrix} = p\begin{pmatrix} p-1 \\ q-1 \end{pmatrix}.$

LEMMA 1.2. $\begin{pmatrix} p \\ r \end{pmatrix} = \begin{pmatrix} p \\ p-r \end{pmatrix}.$

LEMMA 1.3. $\sum_r \begin{pmatrix} p \\ r \end{pmatrix}\begin{pmatrix} p \\ r+s \end{pmatrix} = \begin{pmatrix} 2p \\ p+s \end{pmatrix}.$

PROOF:

$$\sum_s \begin{pmatrix} 2p \\ s \end{pmatrix} x^s = (1+x)^{2p} = \{(1+x)^p\}^2 = \left\{ \sum_s \begin{pmatrix} p \\ s \end{pmatrix} x^s \right\}^2$$

$$= \sum_s \sum_r \begin{pmatrix} p \\ r \end{pmatrix}\begin{pmatrix} p \\ s-r \end{pmatrix} x^s.$$

Hence

$$\begin{pmatrix} 2p \\ s \end{pmatrix} = \sum_r \begin{pmatrix} p \\ r \end{pmatrix}\begin{pmatrix} p \\ s-r \end{pmatrix},$$

and

$$\begin{pmatrix} 2p \\ p+s \end{pmatrix} = \sum_r \begin{pmatrix} p \\ r \end{pmatrix}\begin{pmatrix} p \\ p+s-r \end{pmatrix} = \sum_r \begin{pmatrix} p \\ r \end{pmatrix}\begin{pmatrix} p \\ r-s \end{pmatrix} = \sum_r \begin{pmatrix} p \\ r+s \end{pmatrix}\begin{pmatrix} p \\ r \end{pmatrix}.$$

LEMMA 1.4.   If $p^2 + r^2 > 0$ then $\begin{pmatrix} p \\ r \end{pmatrix} = \begin{pmatrix} p-1 \\ r \end{pmatrix} + \begin{pmatrix} p-1 \\ r-1 \end{pmatrix}.$

LEMMA 1.5. $(p-2r)\begin{pmatrix} p \\ r \end{pmatrix} = p\left\{ \begin{pmatrix} p-1 \\ r \end{pmatrix} - \begin{pmatrix} p-1 \\ r-1 \end{pmatrix} \right\}.$

LEMMA 1.6. $(p-2r)\begin{pmatrix} p \\ r \end{pmatrix}^2 = p\left\{ \begin{pmatrix} p-1 \\ r \end{pmatrix}^2 - \begin{pmatrix} p-1 \\ r-1 \end{pmatrix}^2 \right\}.$

PROOF: Multiply, using 1.4 and 1.5.

LEMMA 1.7.[3] $r\begin{pmatrix} 2p \\ p+r \end{pmatrix}^2 = p\left\{ \begin{pmatrix} 2p-1 \\ p-r \end{pmatrix}^2 - \begin{pmatrix} 2p-1 \\ p-r-1 \end{pmatrix}^2 \right\}.$

---

[3] Major A. A. Bennett communicated this Lemma.

PROOF: $(s - 2t)\binom{s}{t}^2 = s\left\{\binom{s-1}{t}^2 - \binom{s-1}{t-1}^2\right\}$ from 1.6.

Put $s = 2p$, $t = p - r$, then

$$2r\binom{2p}{p+r}^2 = 2p\left\{\binom{2p-1}{p-r}^2 - \binom{2p-1}{p-r-1}^2\right\}.$$

LEMMA 1.8.   *If f is a function, i, n, p are integers and $p + 1 \le i \le n$, then*

$$\sum_{r=1}^{n}\binom{p}{i-r}f(i-r) = \sum_{r=0}^{p}\binom{p}{r}f(r).$$

PROOF:

$$\sum_{r=1}^{n}\binom{p}{i-r}f(i-r) = \sum_{s=i-n}^{i-1}\binom{p}{s}f(s) = \sum_{r=0}^{p}\binom{p}{r}f(r).$$

LEMMA 1.9.   *If $-\infty < A(r, s) = A(s, r) < \infty$ for each integer r and s, then*

$$E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}A(r,s)\epsilon_r\epsilon_s\right\}^2\right) = (\mu_4 - 3\sigma^4)\sum_{r=1}^{n}A(r,r)^2$$

$$+ \sigma^4\left\{\sum_{r=1}^{n}A(r,r)\right\}^2 + 2\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}A(r,s)^2.$$

PROOF: Let $N(r, s) = 1$ when $r < s$ and let $N(r, s) = 0$ otherwise.   Clearly

$$\sum_{r=1}^{n}\sum_{s=1}^{n}A(r,s)\epsilon_r\epsilon_s = \sum_{r=1}^{n}A(r,r)\epsilon_r^2 + 2\sum_{r=1}^{n}\sum_{s=1}^{n}N(r,s)A(r,s)\epsilon_r\epsilon_s,$$

and

$$E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}A(r,s)\epsilon_r\epsilon_s\right\}^2\right) = E\left(\left\{\sum_{r=1}^{n}A(r,r)\epsilon_r^2\right\}^2\right)$$

$$+ 4E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}N(r,s)A(r,s)\epsilon_r\epsilon_s\right\}^2\right).$$

Now

$$E\left(\left\{\sum_{r=1}^{n}A(r,r)\epsilon_r^2\right\}^2\right) = (\mu_4 - \sigma^4)\sum_{r=1}^{n}A(r,r)^2 + \sigma^4\left\{\sum_{r=1}^{n}A(r,r)\right\}^2,$$

and

$$4E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}N(r,s)A(r,s)\epsilon_r\epsilon_s\right\}^2\right)$$

$$= 4\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}N(r,s)A(r,s)^2$$

$$= 2\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}A(r,s)^2 - 2\sigma^4\sum_{r=1}^{n}A(r,r)^2$$

The last three relations combine to yield the desired result.

LEMMA 1.10. $\dbinom{2p}{p}^2 (n-p)^2 E(\delta^4_{n,p})$

$$= (\mu_4 - 3\sigma^4) \sum_{r=1}^{n} \left\{ \sum_{s=1}^{n} \dbinom{p}{i-r}^2 \right\}^2 + \sigma^4 \left\{ \sum_{r=1}^{n} \sum_{s=1}^{n} \dbinom{p}{i-r}^2 \right\}^2$$

$$+ 2\sigma^4 \sum_{r=1}^{n} \sum_{s=1}^{n} \left\{ \sum_{i=p+1}^{n} \dbinom{p}{i-r} \dbinom{p}{i-s} \right\}^2 .$$

PROOF: Helped by 1.8, check that

$$(\Delta_i^p \epsilon)^2 = \left\{ \sum_{r=0}^{p} (-1)^r \dbinom{p}{r} \epsilon_{i-r} \right\}^2 = \left\{ \sum_{r=1}^{n} (-1)^{i-r} \dbinom{p}{i-r} \epsilon_r \right\}^2$$

$$= \sum_{r=1}^{n} \sum_{s=1}^{n} (-1)^{r+s} \dbinom{p}{i-r} \dbinom{p}{i-s} \epsilon_r \epsilon_s .$$

Therefore

$$\dbinom{2p}{p} (n-p)\delta^2_{n,p} = \sum_{r=1}^{n} \sum_{s=1}^{n} \left\{ (-1)^{r+s} \sum_{i=p+1}^{n} \dbinom{p}{i-r} \dbinom{p}{i-s} \right\} \epsilon_r \epsilon_s .$$

Let

$$A(r,s) = (-1)^{r+s} \sum_{i=p+1}^{n} \dbinom{p}{i-r} \dbinom{p}{i-s} ,$$

and apply 1.9 to complete the proof.

LEMMA 1.11.

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \left\{ \sum_{i=p+1}^{n} \dbinom{p}{i-r} \dbinom{p}{i-s} \right\}^2$$

$$= (n-p) \sum_{r=p-n}^{n-p} \dbinom{2p}{p+r}^2 - 2p \dbinom{2p-1}{p}^2 + 2p \dbinom{2p-1}{n}^2$$

PROOF.

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \left\{ \sum_{i=p+1}^{n} \dbinom{p}{i-r} \dbinom{p}{i-s} \right\}^2$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \sum_{r=1}^{n} \sum_{s=1}^{n} \dbinom{p}{i-s} \dbinom{p}{j-s} \dbinom{p}{i-r} \dbinom{p}{j-r}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \sum_{r=1}^{n} \sum_{s=0}^{p} \dbinom{p}{s} \dbinom{p}{s+j-i} \dbinom{p}{i-r} \dbinom{p}{j-r} , \text{ using 1.8;}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \sum_{r=0}^{p} \sum_{s=0}^{p} \dbinom{p}{s} \dbinom{p}{s+j-i} \dbinom{p}{r} \dbinom{p}{r+j-i} , \text{ using 1.8 again;}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \sum_{r} \binom{p}{r} \binom{p}{r+j-i} \sum_{s} \binom{p}{s} \binom{p}{s+j-i}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \left\{ \sum_{r} \binom{p}{r} \binom{p}{r+j-i} \right\}^2 = \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \binom{2p}{p+j-i}^2, \text{ from 1.3;}$$

$$= \sum_{r=p+1-n}^{n-p-1} (n-p-|r|) \binom{2p}{p+r}^2$$

$$= \sum_{r=p-n}^{n-p} (n-p-|r|) \binom{2p}{p+r}^2$$

$$= (n-p) \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2 \sum_{r=0}^{n-p} r \binom{2p}{p+r}^2$$

$$= (n-p) \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2 \sum_{r=0}^{n-p} p \left\{ \binom{2p-1}{p-r}^2 - \binom{2p-1}{p-r-1}^2 \right\}, \text{ using 1.7;}$$

$$= (n-p) \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2p \left\{ \binom{2p-1}{p}^2 - \binom{2p-1}{2p-n-1}^2 \right\}$$

$$= (n-p) \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2p \binom{2p-1}{p}^2 + 2p \binom{2p-1}{n}^2.$$

LEMMA 1.12.

$$\sum_{r=1}^{n} \sum_{i=p+1}^{n} \binom{p}{i-r}^2 = (n-p) \binom{2p}{p}.$$

PROOF.

$$\sum_{r=1}^{n} \sum_{i=p+1}^{n} \binom{p}{i-r}^2 = \sum_{i=p+1}^{n} \sum_{r=1}^{n} \binom{p}{i-r}^2 = \sum_{i=p+1}^{n} \sum_{r=0}^{p} \binom{p}{r}^2, \text{ from 1.8;}$$

$$= (n-p) \sum_{r} \binom{p}{r}^2 = (n-p) \binom{2p}{p}$$

7. The variances of $\delta_{n,p}^2$ and $d_{n,p}^2$. In order to get some idea as to the efficiency of the statistics $\delta_{n,p}^2$ and $d_{n,p}^2$, we will examine their variances. We have

$$\binom{2p}{p}^2 (n-p)^2 \text{Var}(\delta_{n,p}^2) = \binom{2p}{p}^2 (n-p)^2 \{ E(\delta_{n,p}^4) - [E(\delta_{n,p}^2)]^2 \}$$

$$= \binom{2p}{p}^2 (n-p)^2 \sigma^4 + 2(n-p)\sigma^4 \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 4p\sigma^4 \binom{2p-1}{p}^2$$

$$+ 4p\sigma^4 \binom{2p-1}{n}^2$$

with the aid of Lemmas 1.10, 1.11, 1.12 and using the relation $\mu_4 - 3\sigma^4 = 0$.
  Thus,

$$
\begin{aligned}
(8) \quad &\binom{2p}{p}^2 (n-p)^2 \operatorname{Var}(\delta^2_{n,p}) \\
&= 2(n-p)\sigma^4 \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 4p\sigma^4 \binom{2p-1}{p}^2 + 4p\sigma^4 \binom{2p-1}{n}^2.
\end{aligned}
$$

If $2p \leq n$, then

$$
\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 \sum_r \binom{2p}{p+r}^2 = \sum_r \binom{2p}{r} = \binom{4p}{2p}.
$$

Moreover, $\binom{2p-1}{n} = 0$.

Therefore,

$$
(9) \quad \binom{2p}{p}^2 (n-p)^2 \operatorname{Var}(\delta^2_{n,p}) = 2(n-p)\binom{4p}{2p}\sigma^4 - 4p\binom{2p-1}{p}^2 \sigma^4
$$

when $2p \leq n$.
As for the variance of $d^2_{n,p}$, we have

$$
\begin{aligned}
\operatorname{Var}(d^2_{n,p}) &= E\{d^2_{n,p} - \nu^2_{n,p} - \sigma^2\}^2 = E\{\delta^2_{n,p} + k_{n,p} + \nu^2_{n,p} - \nu^2_{n,p} - \sigma^2\}^2 \\
&= E\{\delta^2_{n,p} - \sigma^2) + k_{n,p}\}^2,
\end{aligned}
$$

or

$$
(10) \quad \operatorname{Var}(d^2_{n,p}) = \operatorname{Var}(\delta_{n,p})^2 + E(k^2_{n,p}),
$$

since $E[(\delta^2_{n,p} - \sigma^2)k_{n,p}] = 0$.
However, from Schwarz's inequality, it is guaranteed that

$$
E(k^2_{n,p}) \leq 4\nu^2_{n,p}\sigma^2.
$$

Thus

$$
(11) \quad \operatorname{Var} d^2_{n,p} \leq \operatorname{Var}(\delta^2_{n,p}) + 4\nu^2_{n,p}\sigma^2.
$$

An upper bound has already been given for $\nu^2_{n,p}$ in section 5 above.

8. The efficiency of $\delta^2_{n,p}$. It is appropriate to consider the efficiency (as defined by Fisher [11]) of the statistic $\delta^2_{n,p}$. In this sense, the efficiency of $\delta^2_{n,p}$ is given by

$$
W(n, p) = \frac{\operatorname{Var} s^2_n}{\operatorname{Var} \delta^2_{n,p}}. \qquad \text{where } s^2_n = \frac{\sum (x_i - \bar{x})^2}{n-1}.
$$

Accordingly,

$$W(n, p) = \frac{2\sigma^4}{(n-1)\,\mathrm{Var}\,(\delta^2_{n,p})}$$

or

$W(n, p)$

(12)

$$= \frac{(n-p)^2 \binom{2p}{p}^2}{(n-1)\left\{(n-p)\sum_{r=p-n}^{n-p}\binom{2p}{p+r}^2 - 2p\binom{2p-1}{p}^2 + 2p\binom{2p-1}{n}^2\right\}}.$$

If $2p \leq n$

(13) $\qquad W(n.\ p) = \dfrac{(n-p)^2 \binom{2p}{p}^2}{(n-1)\left\{(n-p)\binom{4p}{2p} - 2p\binom{2p-1}{p}^2\right\}}, \qquad$ from (9);

or

(14) $\qquad W(n, p) = \dfrac{\binom{2p}{p}^2}{\binom{4p}{2p}\left\{1 - \dfrac{p-1}{n-p}\right\}\left\{1 - \dfrac{2p\binom{2p-1}{p}^2}{(n-p)\binom{4p}{2p}}\right\}}, \qquad$ if $2p \leq n.$

Formulas (12) and (13) were used in preparing Table I given at the end of the paper. For convenience in using formulas (1) and (2) the binomial coefficients $\binom{2p}{p}$ for $0 \leq p \leq 10$ are given in Table III.

If $n \geq 2$, then

(15) $\qquad W(n, 1) = \dfrac{2}{3} \cdot \dfrac{1}{1 - \dfrac{1}{3n-3}} = \dfrac{2(n-1)}{3n-4},$

as was pointed out by von Neumann, Kent, Bellinson, and Hart in [3].
If $n \geq 4$, then

(16) $\qquad W(n, 2) = \dfrac{18}{35}\dfrac{1}{\left\{1 + \dfrac{1}{n-2}\right\}\left\{1 - \dfrac{18}{35(n-2)}\right\}} = \dfrac{18(n-2)^2}{(n-1)(35n-88)}.$

As a limiting value for n, we have

(17)
$$W(\infty, p) = \operatorname*{Lim}_{n \to \infty} W(n, p) = \frac{\binom{2p}{p}^2}{\binom{4p}{2p}}.$$

Using Stirling's formula for the approximation to the factorial, we have

$$\operatorname*{Lim}_{p \to \infty} \sqrt{p} W(\infty, p) = \sqrt{\frac{2}{\pi}}.$$

Thus, as $p \to \infty$, $W(\infty, p)$ tends to zero and is asymptotically equal to $\sqrt{\dfrac{2}{\pi p}}$

TABLE III

The Binomial Coefficient $\binom{2p}{p}$

p	$\binom{2p}{p}$
0	1
1	2
2	6
3	20
4	70
5	252
6	924
7	3432
8	12870
9	48620
10	184756

For the case $n \geq 2$, $p \geq 1$ and f *constant*, then $s_n^2 = \dfrac{\Sigma (x_i - \bar{x})^2}{n - 1}$ and $\delta_{n,p}^2$ and $d_{n,p}^2$ are all unbiased estimates of the population variance σ^2. Moreover, for this case

$$W(n, p) = \frac{\operatorname{Var}(s_n^2)}{\operatorname{Var}(\delta_{n,p}^2)} = \frac{\operatorname{Var}(s_n^2)}{\operatorname{Var}(d_{n,p}^2)}.$$

Using s_m^2 based on $m - 1$ degrees of freedom and keeping the trend, f, constant, then m and n may be chosen so that approximately

$$\operatorname{Var}(s_m^2) = \operatorname{Var}(d_{n,p}^2)$$

and for a normal population this means that

$$m = 1 + (n - 1)W(n, p).$$

Using Table I, it may be seen that for constant trend, f, the worth of $d_{50,10}^2$ as an estimate of σ^2 for a normal population is about the same as that of s_{11}^2, whereas that of $d_{50,1}^2$ is about equivalent to s_{40}^2. However, if the trend f is not constant then the worth of s_n^2 as an estimate of σ^2 is diminished while that of $d_{n,p}^2$ is increased.

Similarly, if the trend is cubic over 20 observations then least squares gives an unbiased estimate of σ^2 based on 16 degrees of freedom, whereas $d_{20,4}^2$ gives an estimate equivalent in precision to about 6.4 degrees of freedom. However, if only eight observations follow a cubic trend, then least squares furnish an unbiased estimate of σ^2 based on four degrees of freedom whereas $d_{8,4}^2$ furnishes an estimate equivalent to about 1.9 degrees of freedom. Thus, in the case of 20 observations, cubic least squares is, so to speak, 2.5 times as valuable as $d_{20,4}^2$; in the case of eight observations, cubic least squares is 2.1 times as valuable as $d_{8,4}^2$.

It might be mentioned that the method of differences is of value in estimating goodness of fit. If the fit is good, then our estimate of σ^2 derived from least squares should on the average be equal to the estimate derived from a suitable $d_{n,p}^2$. If the fit is poor then $d_{n,p}^2$ will be smaller on the average than the former.

9. The approximate probable error in estimating σ from differences. The approximate standard error of $\delta_{n,p}$ is given by the relation

$$\text{S.E. } (\delta_{n,p}) \sim \frac{1}{2} \frac{\text{S.E. } (\delta_{n,p}^2)}{\sigma} = \frac{\sigma}{\sqrt{2(n-1)W(n,p)}}.$$

If p has been so chosen that $\nu_{n,p}^2$ is suitably small then [see equation (11)] some confidence may be put in the approximate formulas:

$$(18) \qquad \text{S.E } (d_{n,p}) = \frac{\sigma}{\sqrt{2(n-1)W(n,p)}}$$

$$(19) \qquad \text{P.E. } (d_{n,p}) = \frac{.6745\sigma}{\sqrt{2(n-1)W(n,p)}}.$$

Formula (19) was used in preparing Table IV which gives the approximate probable error to be feared in using $d_{n,p}$ as an estimate of σ. This table should yield interesting information whenever p has been chosen so that $d_{n,p}^2$ is a suitably unbiased estimate of σ^2.

10. Remarks. We have presented a useful technique for estimating variance from higher order differences and have given the precision of our estimate. The method of estimating variance from higher order differences appears to be quite valuable in cases where the type of trend in our observations is unknown. A considerable field of work remains concerning a complete investigation of the distribution and other properties of the statistic $d_{n,p}^2$. In this connection, Baer [12] has already published a study on the stochastic limit of $\frac{n}{n-1} d_{n,1}^2$.

It is hoped that others will contribute to the problem of estimating dispersion

TABLE IV
The Probable Error In Estimating σ From Differences*

n \ p	0	1	2	3	4	5	6	7	8	9	10
1	.4769										
2	.3373	.4769									
3	.2753	.3771	.4769								
4	.2384	.3180	.4054	.4769							
5	.2133	.2796	.3495	.4215	.4769						
6	.1948	.2524	.3104	.3704	.4404	.4769					
7	.1803	.2317	.2817	.3318	.3855	.4390	.4769				
8	.1686	.2154	.2596	.3024	.3477	.3969	.4442	.4769			
9	.1589	.2022	.2420	.2794	.3183	.3604	.4057	.4481	.4769		
10	.1508	.1911	.2274	.2610	.2948	.3311	.3708	.4128	.4513	.4769	
11	.1438	.1816	.2153	.2457	.2758	.3074	.3417	.3794	.4186	.4537	.4769
12	.1376	.1734	.2048	.2328	.2599	.2880	.3180	.3508	.3867	.4234	.4558
13	.1323	.1663	.1958	.2217	.2465	.2717	.2983	.3272	.3587	.3930	.4276
14	.1274	.1599	.1878	.2120	.2350	.2579	.2818	.3073	.3351	.3656	.3984
15	.1231	.1542	.1808	.2035	.2248	.2459	.2677	.2905	.3152	.3423	.3718
16	.1192	.1491	.1744	.1960	.2159	.2355	.2554	.2761	.2983	.3223	.3485
17	.1156	.1445	.1687	.1892	.2080	.2262	.2447	.2637	.2837	.3052	.3286
18	.1124	.1403	.1636	.1831	.2009	.2180	.2352	.2527	.2710	.2905	.3116
19	.1094	.1364	.1589	.1775	.1945	.2106	.2267	.2430	.2599	.2777	.2967
20	.1066	.1328	.1545	.1724	.1886	.2040	.2191	.2343	.2500	.2663	.2837
21	.1040	.1295	.1505	.1677	.1832	.1978	.2121	.2264	.2411	.2562	.2722
22	.1016	.1265	.1468	.1634	.1783	.1922	.2058	.2193	.2331	.2472	.2620
23	.0994	.1236	.1433	.1594	.1738	.1871	.2000	.2129	.2258	.2391	.2529
24	.0973	.1209	.1401	.1557	.1695	.1824	.1948	.2069	.2191	.2316	.2446
25	.0954	.1184	.1371	.1522	.1656	.1779	.1898	.2015	.2131	.2249	.2370
26	.0935	.1160	.1343	.1490	.1619	.1739	.1853	.1964	.2075	.2187	.2301
27	.0918	.1138	.1316	.1459	.1585	.1700	.1810	.1917	.2023	.2130	.2238
28	.0902	.1117	.1291	.1431	.1553	.1664	.1770	.1873	.1975	.2077	.2180
29	.0885	.1097	.1268	.1404	.1522	.1631	.1733	1832	.1930	.2028	.2126
30	.0871	.1078	.1245	.1378	.1493	.1599	.1698	.1794	.1888	.1981	.2076
31	.0857	.1060	.1224	.1354	.1466	.1569	.1665	.1758	.1848	.1938	.2029
32	.0843	.1043	.1204	.1331	.1441	.1540	.1634	.1724	.1811	.1898	.1985
33	.0831	.1027	.1184	.1309	.1416	.1514	.1605	.1692	.1776	.1860	.1944
34	.0818	.1012	.1166	.1288	.1393	.1488	.1577	.1661	.1744	.1825	.1905
35	.0807	.0999	.1149	.1268	.1371	.1464	.1550	.1632	.1713	.1791	.1869
36	.0795	.0983	.1132	.1249	.1350	.1441	.1525	.1605	.1683	.1759	.1834
37	.0784	.0969	.1116	.1231	.1330	.1418	.1501	.1579	.1655	.1729	.1802
38	.0774	.0956	.1101	.1214	.1311	.1397	.1478	.1555	.1628	.1700	.1771
39	.0764	.0943	.1086	.1197	.1292	.1377	.1456	.1531	.1603	.1673	.1741
40	.0754	.0931	.1072	.1181	.1274	.1358	.1435	.1508	.1578	.1646	.1713

TABLE IV—*Continued*

n \ p	0	1	2	3	4	5	6	7	8	9	10
42	.0736	.0909	.1045	.1151	.1241	.1322	.1396	.1466	.1533	.1597	.1661
44	.0719	.0887	.1020	.1123	.1211	.1288	.1360	.1427	.1491	.1553	.1613
46	.0703	.0868	.0997	.1097	.1182	.1257	.1326	.1391	.1453	.1512	.1570
48	.0689	.0849	.0975	.1073	.1155	.1228	.1295	.1357	.1417	.1474	.1529
50	.0675	.0832	.0955	.1050	.1130	.1201	.1266	.1326	.1383	.1438	.1492
52	.0661	.0815	.0936	.1029	.1107	.1176	.1238	.1297	.1352	.1405	.1457
54	.0649	.0800	.0918	.1009	.1085	.1152	.1213	.1270	.1323	.1375	.1425
56	.0637	.0785	.0901	.0990	.1064	.1129	.1189	.1244	.1296	.1346	.1394
58	.0626	.0771	.0885	.0972	.1045	.1108	.1166	.1220	.1271	.1319	.1366
62	.0606	.0746	.0855	.0939	.1008	.1069	.1125	.1176	.1224	.1270	.1313
66	.0587	.0723	.0828	.0909	.0975	.1034	.1087	.1136	.1182	.1225	.1266
70	.0570	.0702	.0804	.0881	.0946	.1002	.1053	.1100	.1144	.1185	.1224
74	.0554	.0682	.0781	.0856	.0919	.0973	.1022	.1067	.1109	.1149	.1186
78	.0540	.0664	.0760	.0833	.0894	.0947	.0994	.1037	.1077	.1115	.1152
82	.0527	.0648	.0741	.0812	.0871	.0922	.0968	.1009	.1048	.1085	.1120
90	.0503	.0618	.0707	.0774	.0830	.0878	.0921	.0960	.0997	.1031	.1063
98	.0482	.0592	.0677	.0741	.0794	.0840	.0880	.0917	.0952	.0984	.1014
106	.0463	.0569	.0650	.0712	.0762	.0806	.0845	.0880	.0913	.0943	.0972
114	.0447	.0549	.0627	.0686	.0734	.0776	.0813	.0847	.0878	.0907	.0934
122	.0432	.0530	.0606	.0663	.0709	.0749	.0785	.0817	.0847	.0875	.0900
138	.0406	.0498	.0569	.0622	.0666	.0703	.0736	.0766	.0794	.0819	.0843
154	.0384	.0472	.0538	.0589	.0630	.0664	.0695	.0723	.0749	.0773	.0795
170	.0366	.0449	.0512	.0560	.0599	.0632	.0661	.0687	.0711	.0734	.0755
202	.0336	.0412	.0470	.0513	.0548	.0578	.0605	.0629	.0650	.0671	.0689
234	.0312	.0382	.0436	.0476	.0509	.0537	.0561	.0583	.0603	.0621	.0639
266	.0292	.0359	.0409	.0446	.0477	.0503	.0525	.0546	.0565	.0582	.0598
330	.0262	.0322	.0367	.0400	.0428	.0451	.0471	.0489	.0505	.0521	.0535
394	.0240	.0295	.0336	.0366	.0391	.0412	.0430	.0447	.0462	.0475	.0488
522	.0209	.0256	.0292	.0318	.0339	.0357	.0373	.0387	.0400	.0412	.0423
778	.0171	.0210	.0239	.0260	.0278	.0292	.0305	.0317	.0327	.0337	.0346
1290	.0133	.0163	.0185	.0202	.0216	.0227	.0237	.0246	.0254	.0261	.0268
2314	.0099	.0121	.0138	.0151	.0161	.0169	.0177	.0183	.0189	.0195	.0200

* If $d_{n.p}^2$ is a sufficiently unbiased estimate of σ^2, then the approximate probable error to be feared in using $d_{n.p}$ as an estimate of σ may be obtained by multiplying the following tabular entries by σ.

when observed data display trends as it is believed that the method of differences deserves much attention. In particular, it is hoped that someone will have the time and ingenuity to calculate the distribution of the statistic

$$\frac{\delta^2_{n,p}}{\delta^2_{n,p+1}}.$$ [4]

Were this done, an admirable criterion would be at hand for gauging the significance of a change in the estimate of σ^2 as we pass from differences of order p to those of order $p + 1$. Of course, useful information in this connection could be obtained from a knowledge of the distributions of $\delta^2_{n,p}$ and $\delta^2_{n,p+1}$; in fact their variances as herein calculated give us a basis for somewhat reasonable conclusions. An expression for the standard error of the difference between the estimates of σ^2 from two consecutive series of finite differences is given in [13, Chapter VI].

In connection with testing goodness of fit, it would be valuable also to know the distribution of

$$\frac{S^2_{n,p}}{\delta^2_{n,p+1}},$$

where $S^2_{n,p}$ is the estimate of variance derived from the least squares fitting of a polynomial of degree p.

For convenience of reference, we conclude the paper with

11. A concise description of the method and its precision. It frequently happens that successive observations made at regular intervals are subject to the same standard error σ while the means of the populations from which they are drawn display a trend.   We give here a method of estimating the variance σ^2 and of determining the precision of our estimate.   This method is primarily of value when the trend is unknown; however even when the type of trend is known, its computational simplicity may make the method advantageous.

The method.   Arrange the data in a vertical column and then in the usual way form difference columns of order $1, 2, \cdots, p$.   Sum the squares of the pth order differences and divide by the number $(n - p)\binom{2p}{p}$.   Our estimate of σ^2 is the number $d^2_{n,p}$, where

$$d^2_{n,p} = \frac{1}{\binom{2p}{p}(n - p)} \sum_{i=p+1}^{n} (\Delta^p x_i)^2.$$

---

[4] Dixon [9] gives moments of the statistic $\dfrac{\sum_{i=1}^{n} (x_i - 2x_{i+1} + x_{i+2})^2}{\sum_{i=1}^{n} (x_i - x_{i+1})^2}$ where $x_{n+1} = x_1$

and $x_{n+2} = x_2$.

The precision. The precision of this estimate may be determined from the following information (which has been derived in the present paper):

$$E(d_{n,p}^2) = \sigma^2 + \nu_{n,p}^2 \;;$$

$$\nu_{n,p}^2 \leq \frac{1}{\binom{2p}{p}} \left(\frac{b-a}{n-p}\right)\left(\frac{b-a}{n-1}\right)^{2p-1} \int_a^b \frac{[f^{(p)}(s)]^2 \, ds}{b-a} \;;$$

$$\mathrm{Var}\,(d_{n,p}^2) \leq \mathrm{Var}\,(\delta_{n,p}^2) + 4\nu_{n,p}^2\sigma^2 \;;$$

$$\mathrm{Var}\,(\delta_{n,p}^2) = \frac{2\sigma^4}{(n-1)W(n,p)},$$

where $W(n, p)$ is given in Table I.

TABLE V

p	σ_x	σ_y	σ_z
1	18.90	184.62	11.22
2	1.21	1.88	10.56
3	.88	1.85	10.30
4	.87	1.84	10.12
5	.86	1.83	10.01

In case $\nu_{n,p}^2$ is sufficiently small (this is determined by the requirements of the given problem), then Table IV may be used directly to determine the approximate probable error in using $d_{n,p}$ as an estimate of σ.

An example. As a practical example of the use of the method of differences when the trend is unknown and of the stability of the statistic $d_{n,p}^2$ with respect to p, we mention a recent problem at Aberdeen Proving Ground which had to do with estimating the accuracy with which certain photographic measurements locate a moving object. Ballistic Cameras were used to determine horizontal x and y, and vertical z coordinates (all in feet) of an airplane traveling about 160 mph at an elevation of about 35,000 feet. An automatic pilot was in use in the airplane as it flew over a three mile course. At one second intervals for a period of 70 seconds two Ballistic Cameras, 5000 feet apart, were used to locate the plane. Since the plane was traveling pretty much in the y direction one would expect: that first differences would yield a standard error in y far in excess of its true one; that second differences would furnish a much better estimate; and that perhaps third differences would yield a still more trustworthy one. No matter what order of difference is used we never expect such an estimate to be too small. In this problem, the standard errors in x, y, z as estimated from differences of certain orders, p, were as given in Table V.

REFERENCES

[1] A. A. BENNETT, unpublished report to the Chief of Ordnance, U. S. Army, circa 1918.

[2] R. H. KENT AND J. VON NEUMANN, "The estimation of the probable error from successive differences", Ballistic Res. Lab. Report No. 175, Feb., 1940.

[3] JOHN VON NEUMANN, R. H. KENT, H. R. BELLINSON, AND B. I. HART, "The mean square successive difference", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 153–162.

[4] J. D. WILLIAMS, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 239–241.

[5] JOHN VON NEUMANN, "Distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 367–395.

[6] JOHN VON NEUMANN, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 86–88.

[7] B. I. HART, "Tabulation of the probabilities for the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 207–214.

[8] B. I. HART, "Significance levels for the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 445–446.

[9] WILFRID J. DIXON, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[10] ANTHONY P. MORSE, "The estimation of dispersion from differences", Ballistic Res Lab. Report No. 557, July, 1945.

[11] R. A. FISHER, *Phil. Trans. A.*, Vol. 222 (1922), p. 316.

[12] REINHOLD BAER, "Sampling from a changing population", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 348–361.

[13] G. TINTNER, *The variate Difference Method*, (Cowles Commission Monograph No 5), Principia Press, Bloomington, Indiana, 1940.

THE EFFICIENCY OF SEQUENTIAL ESTIMATES AND WALD'S EQUATION FOR SEQUENTIAL PROCESSES

By J. Wolfowitz

Columbia University

1. Summary. Let n successive independent observations be made on the same chance variable whose distribution function $f(x, \theta)$ depends on a single parameter θ. The number n is a chance variable which depends upon the outcomes of successive observations; it is precisely defined in the text below. Let $\theta^*(x_1, \cdots, x_n)$ be an estimate of θ whose bias is $b(\theta)$. Subject to certain regularity conditions stated below, it is proved that

$$\sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f}{\partial \theta}\right)^2\right]^{-1}.$$

When $f(x, \theta)$ is the binomial distribution and θ^* is unbiased the lower bound given here specializes to one first announced by Girshick [3], obtained under no doubt different conditions of regularity. When the chance variable n is a constant the lower bound given above is the same as that obtained in [2], page 480, under different conditions of regularity.[1]

Let the parameter θ consist of l components $\theta_1, \cdots, \theta_l$ for which there are given the respective unbiased estimates $\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$. Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the latter, and $\| \lambda^{ij} \|$ its inverse. The concentration ellipsoid in the space of $(k_1, \cdots, k_l)$ is defined as

$$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2.$$

(This valuable concept is due to Cramér). If a unit mass be uniformly distributed over the concentration ellipsoid, the matrix of its products of inertia will coincide with the covariance matrix $\| \lambda_{ij} \|$. In [4] Cramér proves that no matter what the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$, (provided that certain regularity conditions are fulfilled), when n is constant their concentration ellipsoid always contains within itself the ellipsoid

$$\sum_{i,j} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right).$$

---

[1] To whom this result is to be ascribed is not clear from the context in which Professor Cramér describes it (in [2]). After the present paper was completed the author learned of the papers by Rao [8] and Aitken and Silverstone [9], both of which deal with this question. The author is indebted to Prof. M. S. Bartlett for drawing his attention to these papers.

Consider now the sequential procedure of this paper. Let $\theta_1^*, \cdots, \theta_l^*$ be, as before, unbiased estimates of $\theta_1, \cdots, \theta_l$, respectively, recalling, however, that the number n of observations is a chance variable. It is proved that the concentration ellipsoid of $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

$$\sum_{i,j} \mu_{ij}'(k_i - \theta_i)\,(k_j - \theta_j) = l + 2$$

where

$$\mu_{ij}' = EnE\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right).$$

When n is a constant this becomes Cramér's result (under different conditions of regularity).

In section 7 is presented a number of results related to the equation $EZ_n = EnEX$, which is due to Wald [6] and is fundamental for sequential analysis.

2. Introduction. Let X be a chance variable whose distribution function $f(x, \theta)$ depends on the parameter θ. It is assumed that X either has a probability density function (which we then denote by $f(x, \theta)$) or that it can take only an at most denumerable number of discrete values (in the latter case $f(x, \theta) = P\{X = x\}$, where the latter symbol denotes the probability of the relation in braces). Let $\omega = x_1, x_2, \cdots$ be an infinite sequence of observations on X, and let Ω be the space of "points" ω. Let there be given an infinite sequence of Borel measurable functions $\varphi_1(x_1), \varphi_2(x_1, x_2), \cdots, \varphi_j(x_1, \cdots, x_j), \cdots$ defined for all ω in Ω, such that each takes only the values zero and one. It is well known that the function $f(x, \theta)$ defines a measure (probability) on a Borel field in Ω. We assume that everywhere in Ω, except possibly on a set whose probability is zero for all θ under consideration, at least one of the functions $\varphi_1, \varphi_2, \cdots$ takes the value one. Let $n(\omega)$ be the smallest integer at which this occurs. Thus $n(\omega)$ is a chance variable.

In statistical applications the chance variable $n(\omega)$ may be interpreted as a rule for terminating a sequence of observations on the chance variable X, the probability of termination being one, and the decision to terminate depending only upon the observations obtained. A sequential test is an example of this procedure. The converse is, however, not true, because the process described above does not require that any statistical decision should be reached when the process of drawing observations is terminated.

An "estimate" of θ is a function $\theta^*(x_1, \cdots, x_n)$ of the observations $x_1, \cdots, x_n$ (those obtained prior to the "termination" of the process of drawing observations). In the sequel we shall limit ourselves to estimates whose second moments are finite. The estimate is "unbiased" if $E\theta^*$, the expected value of θ^*, is θ. When this is not so $E\theta^* - \theta$ is called the bias, $b(\theta)$, of θ^*. In general the bias is a function of θ. It is obvious that the function θ^* may be undefined on a set of points $(x_1, \cdots, x_n)$ whose probability is zero for all θ under consideration.

In the present paper we shall be concerned with an upper bound on the efficiency of a sequential estimate, or, more precisely, with a lower bound on its variance. This lower bound is intimately related to certain results on the efficiency of the maximum likelihood estimate from a sample of fixed size. This is not surprising since fixed-size sampling is a special instance of sequential sampling. The results obtained in this paper are also obviously and intimately related to those due to Cramér [4] and those described by him in [2], pp. 477–488. Naturally the conditions of regularity (restrictions on $f(x, \theta)$, θ^*, etc.) under which the results are proved are different. For example, no restrictions on the sequential sampling procedure need appear in the statement of a theorem which deals only with samples of fixed size.

The argument below proceeds as if $f(x, \theta)$ were a probability density function. The results apply equally well to the case where $f(x, \theta)$ is the probability function of a discrete chance variable provided:

1). Integration is replaced by summation wherever this is obviously required.

2). The phrase "almost all points" in a Euclidean space of any finite dimensionality is understood

a). as all points in the space with the possible exception of a set of Lebesgue measure zero, when $f(x, \theta)$ is a probability density function

b). as all points in the space with the possible exception of points one of whose coordinates is a member of the set Z, when $f(x, \theta)$ is the probability function of a discrete chance variable. The set Z consists of all points z such that $f(z, \theta) = 0$ identically for all θ under consideration.

3. Conditions of regularity. In this section we shall formulate the restrictions which we impose on f, the estimates, and the sequential process. They are intended to be such as will be satisfied in most cases of statistical interest. No doubt they can be weakened, but the author has decided against attempting to do so here. The list may seem long for two reasons. Seldom in the literature are the assumptions which, for example, lead to validation of differentiation under the integral sign etc., formulated explicitly. The presence of a sequential procedure means that additional restrictions must be imposed.

In this section we assume that θ is a single parameter. The case where θ has more than one component is treated later.

(3.1). *The parameter θ lies in an open interval D of the real line. D may consist of the entire line or of an entire half-line.*

(3.2). *The derivative $\dfrac{\partial f}{\partial \theta}$ exists for all θ in D and almost all x. We define*

$\dfrac{\partial \log f(x, \theta)}{\partial \theta}$ *as zero whenever $f(x, \theta) = 0$; thus $\dfrac{\partial \log f}{\partial \theta}$ is defined for all θ in D and*

almost all x. We postulate that $E \dfrac{\partial \log f(x, \theta)}{\partial \theta} = 0$ and that $E \left(\dfrac{\partial \log f(x, \theta)}{\partial \theta} \right)^2$

be not zero for all θ in D.

(3.3).
$$E\left(\sum_{i=1}^{n} \left| \frac{\partial \log f(x_i, \theta)}{\partial \theta} \right| \right)^2$$

exists for all θ in D.

(3.4). *Let R_j, $(j = 1, 2, \cdots)$, be the set of points $(x_1, \cdots, x_j)$ in the j-dimensional Euclidean space such that*

$$\varphi_i(x_1, \cdots, x_i) = 0 \qquad\qquad i = 1, 2, \cdots, j-1$$

$$\varphi_j(x_1, \cdots, x_j) = 1.$$

For any integral j there exists a non-negative L-measurable function $T_j(x_1, \cdots, x_j)$ such that

a).
$$\left| \theta^*(x_1, \cdots, x_j) \frac{\partial}{\partial \theta} \prod_{i=1}^{j} f(x_i, \theta) \right| < T_j(x_1, \cdots, x_j)$$

for all θ in D and almost all $(x_1, \cdots, x_j)$ in R_j

b).
$$\int_{R_j} T_j(x_1, \cdots, x_j)\, dx_1 \cdots dx_j$$

is finite.

(3.5). *Let*

$$t_j(\theta) = \int_{R_j} \theta^*(x_1, \cdots, x_j) \prod_{i=1}^{j} f(x_i, \theta)\, dx_i, \qquad\qquad (j = 1, 2, \cdots).$$

We postulate the uniform convergence of the series

$$\sum_{j} \frac{dt_j(\theta)}{d\theta}$$

(the existence of $\dfrac{dt_j(\theta)}{d\theta}$ is a consequence of Assumption (3.4)) for all θ in D.

4. The case of one parameter. In this section we assume that $f(x, \theta)$ depends on a single parameter θ. In sections 5 and 6 we shall discuss the case when θ is a vector with more than one component.

We have $E \dfrac{\partial \log f(x, \theta)}{\partial \theta} = 0$

by (3.2). Define the chance variable

$$Y_n = \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta}.$$

By an argument almost identical with that of [1], Theorem 1, or of Theorem 7.1 below, we have

(4.1) $EY_n = 0.$

From Theorem 7.2 below we obtain

$$(4.2) \qquad \sigma^2(Y_n) = EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2.$$

Let θ^* $(x_1, \cdots, x_n)$ be an estimate of θ such that

$$E\theta^* = \theta + b(\theta).$$

Then

$$(4.3) \qquad \sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, \cdots, x_j) \prod_{i=1}^{j} f(x_i, \theta) \, dx_i = \theta + b(\theta).$$

Differentiation of both members of (4.3) with respect to θ (Assumptions (3.4) and (3.5)) gives

$$(4.4) \qquad E\theta^* \, Y_n = 1 + \frac{db}{d\theta}.$$

From (4.1) it follows that (4.4) gives the covariance between θ^* and Y_n. Hence from (4.2)

$$(4.5) \qquad \sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}.$$

When the bias $b(\theta)$ is constant, for example when $b(\theta) \equiv 0$ in case θ^* is an unbiased estimate, we have from (4.5)

$$(4.6) \qquad \sigma^2(\theta^*) \geq \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}.$$

The equality sign in (4.6) will hold if θ^* may be written as $Z'(\theta)Y_n + Z''(\theta)$, where Z' and Z'' are functions of θ. However, θ^* itself should not be a function of θ if our argument is to remain valid. The subject is connected with the question of the existence of a sufficient estimate.

Let $f(x, \theta)$ be defined as follows:

$$f(x, \theta) = \theta^x(1 - \theta)^{1-x}, \qquad (x = 0 \text{ or } 1; 0 < \theta < 1).$$

Then

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1 - x)}{(1 - \theta)}, \qquad E\left(\frac{\partial \log f}{\partial \theta}\right)^2 = \frac{1}{\theta(1 - \theta)}.$$

Suppose θ^* is unbiased. Then $\sigma^2(\theta^*) \geq \theta(1 - \theta)(En)^{-1}$, a result first given by Girshick [3] under unspecified regularity conditions.

Let the functions $\varphi_1, \varphi_2, \cdots$ be such that $n(\omega)$ is a constant. We are then dealing with samples of fixed size. The result (4.5) is then given in [2], p. 480, under different conditions of regularity.

5. Regularity conditions for the case when θ has more than one component.
We suppose that $\theta = (\theta_1, \cdots, \theta_l)$ and that simultaneous estimates

$\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$ of the components of θ are under discussion. In the sequel we shall limit ourselves to the case when these estimates are all unbiased.

We postulate the following regularity conditions which are sufficient to validate section 6:

(5.1). *The covariance matrix of the estimates* $\theta_1^*, \cdots, \theta_l^*$ *is non-singular for all* θ *in* D (*this time* D *is an open interval of the* l-*dimensional parameter space*).

(5.2). *The conditions of section 3 are satisfied for each* θ_i *and* θ_i^* ($i = 1, \cdots, l$).

6. The ellipsoid of concentration when θ has more than one component. Let

$$\theta = (\theta_1, \cdots, \theta_l).$$

We shall first describe briefly the result of Cramér [4] which refers to samples of fixed size $n > l$. Let $\theta_i^*(x_1, \cdots, x_n)$ be an unbiased estimate of θ_i, ($i = 1, \cdots, l$). Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the θ_i^*, and let $\| \lambda^{ij} \|$ be its inverse. The "ellipsoid of concentration" in the space of points $(k_1, \cdots, k_l)$ is defined as

$$(6.1) \qquad \sum_{i,j=1}^{l} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2.$$

If a unit mass be distributed uniformly over this ellipsoid it will have the point $(\theta_1, \cdots, \theta_l)$ as its center of gravity and λ_{ij} as its product of inertia about the corresponding axes. Cramér proves that, subject to certain regularity conditions, there is a fixed ellipsoid

$$(6.2) \qquad \sum_{i,j=1}^{l} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right)$$

which is always contained entirely within the concentration ellipsoid of any set of unbiased estimates. The two ellipsoids coincide only under certain conditions, among which is that the θ_i^* be jointly sufficient estimates of the θ_i.

Let us now consider the sequential procedure of this paper and postulate the regularity conditions of section 5. Let

$$K = \| k_{ij} \|$$

be a matrix with real elements such that $| K | = 1$ and let

$$K^{-1} = \| k^{ij} \|$$

be its inverse. Let

$$\| \theta \| = \begin{Vmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_l \end{Vmatrix}, \qquad \| \theta^* \| = \begin{Vmatrix} \theta_1^* \\ \cdot \\ \cdot \\ \cdot \\ \theta_l^* \end{Vmatrix}, \qquad \| \psi \| = \begin{Vmatrix} \psi_1 \\ \cdot \\ \cdot \\ \cdot \\ \psi_l \end{Vmatrix}$$

be column matrices. Suppose

(6.3) $$\| \psi \| = K \| \theta \|.$$

Then

(6.4) $$\| \theta \| = K^{-1} \| \psi \|.$$

Define

$$\| \psi^* \| = \left\| \begin{matrix} \psi_1^* \\ \cdot \\ \cdot \\ \cdot \\ \psi_l^* \end{matrix} \right\| = K \| \theta^* \|.$$

From section 4 we have

(6.5) $$EnE \left(\frac{\partial \log f(x, \theta)}{\partial \psi_1} \right)^2 \geq [\sigma^2(\psi_1^*)]^{-1}$$

where the differentiation by which $\frac{\partial \log f}{\partial \psi_1}$ is obtained is performed with $\psi_2, \cdots; \psi_l$
held constant.   Consider the last $(l - 1)$ rows of K as fixed and $(k_{11}, k_{12}, \cdots, k_{1l})$
as free to vary subject only to the restriction that $| K | = 1$.   The left member
of (6.5) is then a fixed quantity, while the right member is a function of the first
row of K.   The inequality (6.5) must remain valid for all admissible
$(k_{11}, \cdots, k_{1l})$.   Hence (6.5) will remain valid if the right member of (6.5) is
replaced by its maximum with respect to $(k_{11}, \cdots, k_{1l})$.   We shall obtain this
maximum and find that (6.5) then implies a result about the minimal ellipsoid
of concentration.

The problem is therefore to minimize $\sigma^2(\psi_1^*)$.   Now

(6.6) $$\sigma^2(\psi_1^*) = \sum_{i,j} \lambda_{ij} k_{1i} k_{1j}.$$

The family of ellipsoids in the space of $(k_{11}, \cdots, k_{1l})$

(6.7) $$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c,$$

where c is a running parameter, has all centers located at the origin.   Let

$$(k_{11}^0, \cdots, k_{1l}^0)$$

be the sought-for maximizing values of $(k_{11}, \cdots, k_{1l})$.   From the definitions of
K and K^{-1} we have

(6.8) $$\sum_i k^{i1} k_{1i} = 1$$

where $(k^{11}, k^{21}, \cdots, k^{l1})$ are constants.   It follows that the minimum value
c_0 of $\sigma^2(\psi_1^*)$ is such that the ellipsoid

(6.9) $$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c_0$$

is tangent to the hyperplane (6.8) at the point $(k_{11}^0, \cdots, k_{1l}^0)$.   Now the tan-
gent plane to (6.9) at this point is given by

(6.10) $$\sum_{i,j} \lambda_{ij} k_{1i}^0 k_{1j} = c_0.$$

From (6.8) and (6.10) we obtain

(6.11) $$c_0 k'^1 = \sum_i k_{1i}^0 \lambda_{ij}, \qquad\qquad (j = 1, \cdots, l).$$

Hence

(6.12) $$c_0 \sum_i \lambda^{ij} k^{i1} = k_{1j}^0, \qquad\qquad (j = 1, \cdots, l)$$

from which

(6.13) $$c_0 \sum_{i,j} \lambda^{ij} k^{i1} k^{j1} = 1.$$

We have

(6.14)
$$\frac{\partial \log f}{\partial \psi_1} = \sum_i k^{i1} \frac{\partial \log f}{\partial \theta_i}$$
$$\left(\frac{\partial \log f}{\partial \psi_1}\right)^2 = \sum_{i,j} k^{i1} k^{j1} \frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}.$$

From (6.5), (6.13), (6.14), and the definition of c_0 we conclude that

(6.15) $$\sum_{i,j} \mu'_{ij} k^{i1} k^{j1} \geq \sum_{i,j} \lambda^{ij} k^{i1} k^{j1}$$

where

(6.16) $$\mu'_{ij} = EnE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right).$$

We may restate (6.15) as follows: The concentration ellipsoid

(6.17) $$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

of the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

(6.18) $$\sum_{i,j} \mu'_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where the μ'_{ij} are defined by (6.16).

The question of the coincidence of the two ellipsoids is connected with the question of the existence of sufficient estimates. It may be difficult to state any general results about the concentration ellipsoid of biased estimates without postulating some relationships among the biases and/or their derivatives.

7. On Wald's equation and related results in sequential analysis. In section 4 we referred to a proof by Blackwell [1] of an equation due to Wald [5] which is fundamental in the Wald theory of sequential tests of statistical hypotheses. Here we shall give a perhaps simpler proof of this equation, and then prove several new and related results of general interest for sequential analysis.

The results of Theorems 7.2 and 7.3 below can be obtained by differentiation of Wald's fundamental identity of sequential analysis ([6], [7]). However, the

conditions under which we obtain these results are less stringent than any so far found sufficient to establish the identity and the validity of differentiating it. Theorem 7.4 and its corollaries refer to sequential processes where the chance variables may have different distributions or even be dependent. In the future we hope to return to the question of finding all central moments of Z_n, the problem of generalizing the fundamental identity, and related questions.

For Theorems 7.1, 7.2, and 7.3 we shall assume a chance variable X whose cumulative distribution function $F(x)$ is subject only to whatever restrictions may be explicitly imposed on it in each theorem. We assume the existence of a general sequential process such as is described above, which is subject only to such restrictions as may be explicitly formulated in each theorem. The sequential process of course defines the chance variable n. Let $x_1, x_2, \cdots$ be successive independent observations on X. We define $Z_n = \sum_{i=1}^{n} x_i$. If $E(X)$ and $\sigma^2(X)$ exist we shall denote them by w and σ^2, respectively.

THEOREM 7.1 (Wald [5], Blackwell [1]). *Suppose w and En exist. Then*

$$(7.1) \qquad E(Z_n - nw) = 0.$$

The following theorem, which is a sort of partial converse of Theorem 7.1, is proved concomitantly with Theorem 7.1:

THEOREM 7.1.1. *If EZ_n exists, and if either $P\{X > 0\} = 0$ or $P\{X < 0\} = 0$, then w and En both exist, and*

$$EZ_n = wEn.$$

Actually the same proof suffices for a somewhat stronger form of Theorem 7.1.1:

THEOREM 7.1.2. *If EZ_n exists, and if*

$$E(X_i \mid n = j) \geq 0 \qquad\qquad (\text{or} \leq 0)$$

for all positive integral j such that $P\{n = j\} \neq 0$, and all $i \leq j$, then w and En both exist, and

$$EZ_n = wEn.$$

THEOREM 7.2. *If $E\left(\sum_{i=1}^{n} \mid x_i - w \mid\right)^2$ exists, then σ^2 and En both exist, and*

$$(7.2) \qquad E(Z_n - nw)^2 = \sigma^2 En.$$

We have

$$
(7.3) \quad
\begin{aligned}
E(Z_n - nw) &= E\left(\sum_{i=1}^{n} (x_i - w)\right) = \sum_{j=1}^{\infty} \int_{R_j} \left(\sum_{i=1}^{j} (x_i - w)\right) \prod_{i=1}^{j} dF(x_i) \\
&= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m=i} dF(x_m).
\end{aligned}
$$

Also

$$(7.4) \qquad \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m-i} dF(x_m) = P\{n \geq j\} E(x_j - w) = 0.$$

Hence

$$(7.5) \qquad \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m-i} dF(x_m) = 0.$$

From this (7.1) follows.

Suppose now that the conditions of Theorem 7.2 are fulfilled. We have

$$
\begin{aligned}
E(Z_n - nw)^2 &= \sum_{j=1}^{\infty} \int_{R_j} \left(\sum_{i=1}^{j} (x_i - w) \right)^2 \prod_{m=1}^{m-j} dF(x_m) \\
(7.6) \qquad &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w)^2 \prod_{m=1}^{m-i} dF(x_m) \\
&\quad + 2 \sum_{j=2}^{\infty} \sum_{s=1}^{j-1} \sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m-i} dF(x_m).
\end{aligned}
$$

Let $s < j$ be any two positive integers. Then

$$(7.7) \qquad \sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m-i} dF(x_m) = 0.$$

Hence

$$(7.8) \qquad \sum_{j=2}^{\infty} \sum_{s=1}^{j-1} \sum_{i=j}^{\infty} \int_{R_i} (x_s - w)(x_j - w) \prod_{m=1}^{m-i} dF(x_m) = 0.$$

In a similar manner we obtain

$$(7.9) \qquad \sum_{i=j}^{\infty} \int_{R_i} (x_j - w)^2 \prod_{m=1}^{m-i} dF(x_m) = \sigma^2 P\{n \geq j\}.$$

From (7.6), (7.8), and (7.9) it therefore follows that

$$(7.10) \qquad E(Z_n - nw)^2 = \sigma^2 \sum_{j=1}^{\infty} P\{n \geq j\} = \sigma^2 \sum_{j=1}^{\infty} jP\{n = j\} = \sigma^2 En$$

which is the desired result.

It remains to prove the validity of rearranging the series in (7.3) and (7.6). First, we have

$$(7.11) \qquad \sum_{i=j}^{\infty} \int_{R_i} |x_j - w| \prod_{m=1}^{m-i} dF(x_m) = P\{n \geq j\} E |X - w|.$$

Hence it follows that

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} |x_j - w| \prod_{m=1}^{m-i} dF(x_m) = \sum_{j=1}^{\infty} P\{ n \geq j\} E | X - w |$$

(7.12)

$$= E | X - w | \sum_{j=1}^{\infty} jP\{n = j\} = E | X - w | En.$$

This justifies the rearrangement of terms in the series in (7.3). Second, the series (7.6) is dominated by the series

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w)^2 \prod_{m=1}^{m-i} dF(x_m)$$

(7.13)

$$+ 2 \sum_{j=2}^{\infty} \sum_{s=1}^{j-1} \sum_{i=j}^{\infty} \int_{R_i} | x_s - w | \cdot | x_j - w | \prod_{m=1}^{m-i} dF(x_m)$$

all of whose terms are positive. The series (7.13) converges because

(7.14) $$E \left(\sum_{i=1}^{n} | x_i - w | \right)^2 < + \infty.$$

Hence the rearrangement of the series (7.6) is valid.

In the sequel we require certain sets $R'_j (j = 1, 2, \cdots)$ which we shall define now. Let R^*_{ij}, $i \leq j$, be the totality of all points $(x_1, \cdots, x_j)$ such that

(7.15) $$(x_1, \cdots, x_i) \epsilon R_i.$$

Let R^j be the j-dimensional Euclidean space. Then

(7.16) $$R'_j = R^j - \sum_{i=1}^{j} R^*_{ij}.$$

We shall now prove:

THEOREM 7.3. *Suppose that* $E \left[\sum_{i=1}^{n} | x_i - w | \right]^3$ *and* $En \left[\sum_{i=1}^{n} | x_i - w | \right]$ *exist.*[2] *Then*

(7.17) $$E(Z_n - nw)^3 = w_3 En + 3\sigma^2 En(Z_n - nw)$$

where

$$w_3 = E(X - w)^3$$

exists.

---

[2] The author has succeeded in proving that the existence of $E \left[\sum_{i=1}^{n} | x_i - w | \right]^3$ implies the existence of $E \left[n \sum_{i=1}^{n} | x_i - w | \right]$. The proof will be published subsequently in connection with other results.

PROOF: We have

$$E(Z_n - nw)^3 = \sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - w) \right]^3 \prod_{m=1}^{j} dF(x_m)$$

$$= \sum_{j=1}^{\infty} \int_{R_j} \sum_{i=1}^{j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m)$$

(7.18)

$$+ 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

$$+ 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m)$$

$$+ 6 \sum_{j=3}^{\infty} \int_{R_j} \sum_{i=3}^{j} \sum_{s=2}^{i-1} \sum_{t=1}^{s-1} (x_t - w)(x_s - w)(x_i - w) \prod_{m=1}^{j} dF(x_m).$$

Considering the first term in the right member of (7.18), it follows that

$$\sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - w)^3 \right] \prod_{m=1}^{j} dF(x_m)$$

(7.19)

$$= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m)$$

$$= \sum_{i=1}^{\infty} w_3 P\{n \geq i\}$$

$$= \sum_{i=1}^{\infty} i w_3 P\{n = i\} = w_3 En.$$

All the rearrangements of terms in the operations involved in the proof of Theorem 7.3 are legitimate because the various series are absolutely convergent.

As for the second term in the right member of (7.18), we have

$$\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

(7.20)

$$= \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m)$$

$$= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \int_{R'_{i-1}} (x_s - w) \prod_{m=1}^{i-1} dF(x_m)$$

$$= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s}^{\infty} \int_{R'_i} (x_s - w) \prod_{m=1}^{i} dF(x_m).$$

We now operate on $En(Z_n - nw)$, and obtain

(7.21)

$$En(Z_n - nw) = \sum_{j=1}^{\infty} \int_{R_j} j \sum_{i=1}^{j} (x_i - w) \prod_{m=1}^{j} dF(x_m)$$

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_j} i(x_i - w) \prod_{m=1}^{i} dF(x_m).$$

We observe that

(7.22)
$$
\sum_{i=j}^{\infty} \int_{R_i} i(x_j - w) \prod_{m=1}^{i} dF(x_m)
$$
$$
= j \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m)
$$
$$
+ \sum_{s=j+1}^{\infty} \sum_{i=s}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m).
$$

To evaluate the left member of (7.22), we proceed as follows: It is easy to see that

(7.23)
$$
\sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m) = 0.
$$

Moreover, when $s > j$,

(7.24)
$$
\sum_{i=s}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{i} dF(x_m) = \int_{R'_{s-1}} (x_j - w) \prod_{m=1}^{s-1} dF(x_m).
$$

Hence

(7.25)
$$
\sum_{i=j}^{\infty} \int_{R_i} i\,(x_j - w) \prod_{m=1}^{i} dF(x_m) = \sum_{s=j}^{\infty} \int_{R'_s} (x_j - w) \prod_{m=1}^{s} dF(x_m).
$$

Therefore

(7.26)
$$
En(Z_n - nw) = \sum_{j=1}^{\infty} \sum_{s=j}^{\infty} \int_{R'_s} (x_j - w) \prod_{m=1}^{s} dF(x_m).
$$

It remains now to consider the third term of the right member of (7.18). We have

(7.27)
$$
\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m).
$$
$$
= \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m).
$$

Now, suppose that in the expression

(7.28)
$$
V_{sij} = \int_{R_j} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m)
$$

where $j \geq i > s$, we integrate with respect to all x_m for which $m \geq i$. Then it is not difficult to see that

(7.29)
$$
\sum_{j=i}^{\infty} V_{sij} = 0
$$

for all s and i such that $1 \leq s < i$. Hence from (7.27)

(7.30)
$$
\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m) = 0.
$$

In a similar way it is shown that the fourth term of the right member of (7.18) is zero.

The desired result (7.17) is a direct consequence of (7.18), (7.19), (7.20), (7.26), and (7.30).

Consider now an infinite sequence of chance variables x_1, x_2, $\cdots$, which need not have the same distribution and which may be dependent (in which case they must satisfy the obvious consistency relationships). We take successive observations on these chance variables and define a sequential process as above, which is subject only to such restrictions as we shall explicitly state. Let Z_n maintain its previous definition.

THEOREM 7.4. *Suppose that*

$$(7.31) \qquad\qquad \nu_i = E(X_i \mid n \geq i)$$

exists for all positive integral i for which $P\{n \geq i\} \neq 0$. *In those cases write*

$$(7.32) \qquad\qquad \nu_i' = E(|\,X_i - \nu_i\,|\,\big|\,n \geq i).$$

Suppose also that the series

$$(7.33) \qquad\qquad \sum_{i=1}^{\infty} (\nu_1' + \cdots + \nu_i')P\{n = i\}$$

converges.   Then

$$(7.34) \qquad\qquad E\left[Z_n - \sum_{i=1}^{n} \nu_i \right] = 0.$$

It is regrettable but unavoidable that the mean values ν_i and ν_1' entering into (7.33) and (7.34) be conditional.   The fundamental reason is that the sequential process may drastically modify the distribution of dependent chance variables, so that their distribution for our purposes can only be considered in conjunction with the sequential process itself.   Consider the following example:

$$P\{X_1 = -1\} = \tfrac{1}{2}, \qquad P\{X_1 = 1\} = \tfrac{1}{2}$$

$$P\{X_2 = -2 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = -1 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = 1 \mid X_1 = 1\} = \tfrac{1}{2}$$

$$P\{\dot{X}_2 = 2 \mid X_1 = 1\} = \tfrac{1}{2}.$$

We have $E(X_2) = 0$. Suppose we define the following sequential process: If $X_1 = -1$, $n = 1$, and if $X_1 = 1$, $n = 2$.   It is then clear that for our purposes X_2 can take no negative values and the fact that $E(X_2) = 0$ is of no use to us.

If, however, the chance variables X_1, X_2, $\cdots$ are independent, this difficulty disappears, and we have the following.

COROLLARY 1 TO THEOREM 7.4. *If the chance variables X_1, X_2, $\cdots$ are independent, we have Theorem 7.4 with $\nu_i = E(X_i)$, and $\nu_i' = E \mid X_i - \nu_i \mid$.*

If further all the X_i have the same distribution, we see that Theorem 7.1 is a special case of Theorem 7.4, since the convergence of the series (7.33) is then a consequence of the existence of w and En. From this argument we see, however, that it is not necessary that all the X_i have the same distribution, and we may write the following generalization of Theorem 7.1:

COROLLARY 2 TO THEOREM 7.4. *Let the X_i be independent with, in general, different distributions. Suppose, however, that all ν_i are equal, and all ν_i' are equal, except perhaps for those i such that $P\{n \geq i\} = 0$. Suppose further that En exists. Then* (7.1) *holds.*

Among possible fields of application of Theorem 7.4 are sequential tests of composite statistical hypotheses, and the random walk of a particle governed by probability distributions which are functions of time and the position of the particle. The extension of this theorem to vector chance variables is straightforward. The extension to higher moments may present difficulties. We hope to return to some of these questions in the future.

PROOF OF THEOREM 7.4. This is very elementary. We have

$$E\left(Z_n - \sum_{i=1}^{n} \nu_i \right) = \sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - \nu_i) \right] dF(x_1, \cdots, x_j)$$

$$(7.35) \qquad = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - \nu_j) \, dF(x_1, \cdots, x_i).$$

$$= \sum_{j=1}^{\infty} P\{n \geq j\} E(X_j - \nu_j \mid n \geq j) = 0.$$

The rearrangement of the series is valid because

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} \mid x_j - \nu_j \mid dF(x_1, \cdots, x_i) = \sum_{j=1}^{\infty} \nu_j' P\{n \geq j\}$$

$$(7.36) \qquad\qquad\qquad = \sum_{j=1}^{\infty} (\nu_1' + \cdots + \nu_j') P\{n = j\}$$

which converges by (7.33).

REFERENCES

[1] DAVID BLACKWELL, "On an equation of Wald," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 84–87.
[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.
[3] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[4] H. CRAMÉR, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85–94.

[5] A. WALD, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[6] A. WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–296.

[7] A. WALD, "Differentiation under the expectation sign of the fundamental identity in sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946).

[8] C. R. RAO, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37, No. 3 (Sept., 1945), pp. 81–91.

[9] A. C. AITKEN AND H. SILVERSTONE "On the estimation of statistical parameters," *Proc. Roy. Soc. Edinburgh*, Vol. 61 (1941), pp. 56-62.

ESTIMATION OF LINEAR FUNCTIONS OF CELL PROPORTIONS

BY JOHN H. SMITH

Bureau of Labor Statistics

Summary. In this article certain contributions are made to the theory of estimating linear functions of cell proportions in connection with the methods of (1) least squares, (2) minimum chi-square, and (3) maximum likelihood. Distinctions among these three methods made by previous writers arise out of (1) confusion concerning theoretical vs. practical weights, (2) neglect of effects of correlation between sampling errors, and (3) disagreement concerning methods of minimization. Throughout the paper the equivalence of these three methods from a practical point of view has been emphasized in order to facilitate the integration and adaptation of existing statistical techniques. To this end:

1. The method of least squares as derived by Gauss in 1821–23 [6, pp. 224–228] in which weights in theory are chosen so as to minimize sampling variances is herein called the ideal method of least squares and the theoretical estimates are called ideal linear estimates. This approach avoids confusion between practical approximations and theoretical exact weights.

2. The ideal method of least squares is applied to uncorrelated linear functions of correlated sample frequencies to determine the appropriate quantity to minimize in order to derive ideal linear estimates in sample-frequency problems. This approach leads to a sum of squares of standardized uncorrelated linear functions of sampling errors in which statistics are to be substituted in numerators.

3. A new elementary method is used to reduce the sum of squares in (2)—before substitution of statistics—to Pearson's expression for chi-square. In this result, obtained without approximation, appropriate substitution of statistics shows that the denominators of chi-square should be treated as constant parameters in the differentiation process in order to minimize chi-square in conformity with the ideal method of least squares. .

4. The ideal method of minimum chi-square, derived in (3) as the sample-frequency form of the ideal method of least squares, yields ideal linear estimates in terms of the unknown parameters in the denominators of chi-square. When these parameters are estimated by successive approximations in such a way as to be consistent with statistics based on them, it is shown that the method of minimum chi-square leads to maximum likelihood statistics.

5. An iterative method which converges to maximum likelihood estimates is developed for the case in which observations are cross-classified and first order totals are known. In comparison with Deming's asymptotically efficient statistics, it is shown that, in a certain sense, maximum likelihood statistics are superior for any given value of n—especially in small samples.

6. The method of proportional distribution of marginal adjustments is de-

veloped. This method yields estimates of expected cell frequencies whose efficiency is 100 per cent when universe cell frequencies are proportional—a condition closely approximated in most practical surveys for which first order totals are available from complete censuses. Whether this favorable condition is satisfied or not, the method yields results which are easy to interpret and it has many computational advantages from the point of view of economy of time and effort.

Throughout the article discussion is confined to the estimation of parameters whose relationships to cell proportions are linear. However, most of the results can be extended to the case of non-linear relationships, the necessary qualifications being similar to those in curve-fitting problems when the function to be fitted is not linear in its parameters. In this case, of course, least squares estimates are not linear estimates. In particular, obvious extensions of the general proofs in sections 5 and 6 make them applicable to the non-linear case. Thus even when relationships are non-linear, it can be shown that the method of minimum chi-square is the sample-frequency form of the method of least squares which leads (by means of appropriate successive approximations) to maximum likelihood statistics in sample-frequency problems. This principle which establishes the equivalence of the methods of least squares, minimum chi-square, and maximum likelihood greatly facilitates the integration and adaptation of existing techniques developed in connection with these important methods of estimation.

1. Introduction. This article deals with problems of statistical estimation in which the parameters to be estimated are cell proportions or linear functions of them. A simple illustration of this type of problem is that of estimating p, the proportion of white men in a population classified by race and sex. Fom a sample of n persons selected at random from such a population, the desired proportion can be estimated by simply taking the sample proportion of white men as an estimate of the corresponding cell proportion in the population or universe. This estimate is unbiased for all possible values of p and its sampling variance is $p(1 - p)/n$—assuming, for simplicity, that sampling is done with replacements. Whether a more accurate unbiased estimate of p can be derived depends on whether or not any other relevant information concerning the cell proportions in the universe is available. For example, it may be known that all of the white portion of the population is composed of married couples so that in the universe the number of white men is exactly equal to the number of white women. This knowledge implies that half the proportion of whites provides an unbiased estimate of p which is far more accurate than the sample proportion of white men. In fact, the sampling variance of half the proportion of whites is equal to $(2p)(1 - 2p)/4n$—less than half the sampling variance of the proportion of white men.

The term *ideal linear estimate* will be used to refer to any statistic which satisfies the criteria of estimation implied by the foregoing discussion—that is, an

ideal linear esimate is any estimate which (1) is a linear function of the sample observations; (2) is recognizable as unbiased by the research worker; and (3) has minimum sampling variance among estimates which have properties (1) and (2). These important criteria of estimation will now be stated in more technical language.

Let n_1, n_2, and n_3 represent the number of (1) white men, (2) white women and (3) non-white persons, respectively, in samples of n persons. Since any linear function with a constant term can be reduced to the homogeneous form by adding an appropriate multiple of the identity

(1.1) $$n_1 + n_2 + n_3 - n \equiv 0,$$

it is possible, without loss of generality, to confine attention to linear estimates of the form

(1.2) $$T = a_1 n_1 + a_2 n_2 + a_3 n_3,$$

which are recognizable as unbiased. In this example, the research worker is assumed to know that the cell proportions in the universe are

(1.3) $$p_1, p_2, p_3 = p, p, 1 - 2p.$$

Hence, absence of bias implies that the expected value of T

(1.4) $$E(T) = a_1 n p_1 + a_2 n p_2 + a_3 n p_3$$
$$= (a_1 + a_2 - 2a_3)np + na_3$$

is identically equal to p; in other words, that

(1.5) $$n(a_1 + a_2 - 2a_3) - 1 = 0,$$

and

$$na_3 = 0.$$

The ideal linear esimate is derived by finding values of a_1, a_2, and a_3 which minimize the sampling variance of T subject to equations (1.5) as side conditions.[1] In this way it can be shown that half the sample proportion of whites is actually the ideal linear estimate of p. For more general problems, the process of minimization of sampling variances with the aid of Lagrange multipliers involves expressions which are complicated algebraically. For this reason it is usually easier to derive ideal linear estimates of parameters which are linear functions of cell proportions by the ideal method of least squares which is presented in section 4.

Like other least squares estimates, an ideal linear estimate of a linear function of cell proportions depends on ideal least squares weights. Since these weights

[1] In this example, it is possible to solve equations (1.5) for a_2 in terms of a_1, drop subscripts, and substitute in the formula for the sampling variance of T to obtain a quadratic in a to be minimized.

are, in general, functions of variances and covariances of sample frequencies, the theoretical connotation of the term "ideal" makes it preferable to other terms such as "optimum" and "best." In this connection it should be emphasized that (1) the sampling variance of linear estimates is insensitive to small errors in estimating ideal weights, and (2) the process of deriving practical approximations to ideal linear estimates automatically provides maximum likelihood estimates of the ideal weights. Thus the estimation of weights is perfectly objective and the best practical approximations to ideal linear estimates are expressed in terms of sample observations. This degree of objectivity is rare in statistical estimation as a brief consideration of regression problems will illustrate.

In ordinary regression problems, the ideal weights are inversely proportional to error variances. It is usually necessary to draw upon past experience to estimate relative weights because satisfactory estimates of error variances are rarely available in terms of sample observations. From the present point of view, the widespread use of equal weights implies the *subjective* "assumption" that all error variances are equal. (Maximum likelihood estimates of regression coefficients require, in addition, the even more subjective assumption of normality.) In spite of these (usually implicit) subjective assumptions, discussions of optimum properties of least squares regression coefficients based on *ideal* weights in terms of *unknown parameters* are highly commendable because (1) sampling variance is not very sensitive to small errors in weights and (2) properties of theoretical ideal linear estimates furnish a simple basis for discussion of the properties of practical statistics based on any reasonably good approximations to the exact ideal weights. In any case, it is important to know what the ideal weights are in terms of unknown parameters because research workers can make better estimates if they know what quantities should be estimated than they could otherwise.

2. Estimation of a single parameter. In sample-frequency problems, least squares weights are rarely given explicitly or even implied by information available to the research worker. Since the hypothetical example used in Section 1 is a trivial special case from. this point of view, a more realistic example is presented in this section. Since the biological interpretation of this problem is presented in detail in all but the first of the many editions of Fisher's well-known book [3] it is sufficient here to consider only the statistical problem. The four cell proportions are

$$(2.1) \qquad p_1, p_2, p_3, p_4 = (2 + \theta)/4, (1 - \theta)/4, (1 - \theta)/4, \theta/4,$$

and the parameter θ is to be estimated from the set of sample frequencies

$$(2.2) \qquad n_1, n_2, n_3, n_4 = 1997, 906, 904, 32,$$

obtained in a sample of $n = 3839$ selected at random from an infinite universe. Fisher considers five different statistics—T_1, T_2, T_3, T_4, and T_5—so it will

be convenient to use the symbol T_6 for the ideal linear estimate. Consider the class of linear unbiased estimates of the form

(2.3)
$$T = a_1 n_1 + a_2 n_2 + a_3 n_3 + a_4 n_4,$$

where absence of bias implies that

(2.4)
$$2a_1 + a_2 + a_3 = 0$$

and

$$a_1 - a_2 - a_3 + a_4 - 4/n = 0.$$

Minimizing the sampling variance of T in equation (2.3) subject to side conditions based on equations (2.4) yields the ideal linear estimate T_6 defined by the equation

(2.5)
$$n(1 + 2\theta)T_6 = 3\theta n_1 - 3\theta n_2 - 3\theta n_3 + (4 - \theta)n_4.$$

The exact sampling variance of T_6,

(2.6)
$$\sigma_6^2 = \frac{2\theta(1 - \theta)(2 + \theta)}{n(1 + 2\theta)}$$

is used by Fisher as the asymptotic sampling variance of any efficient estimate of θ. The exact sampling variance of the ideal linear estimate is especially appropriate as the asymptotic sampling variance of the maximum likelihood estimate T_4 because T_4 is the limit of an iterative process designed to estimate T_6 as closely as possible from sample data by using successive approximations to T_6 for θ in equation (2.5). The limit of this process (which is, of course, only an approximation to T_6) can be obtained by substituting the symbol T_4 for both T_6 and θ in equation (2.5) and solving the resulting quadratic equation which can be reduced to

(2.7)
$$n T_4^2 - (n_1 - 2n_2 - 2n_3 - n_4)T_4 - 2n_4 = 0,$$

an equation which is identical, except for notation, with Fisher's equation of maximum likelihood of which T_4 is the positive solution.

The foregoing result is a comparatively simple illustration of the general principle that the maximum likelihood estimate of any linear function of cell proportions is the limit of an iterative process designed to approximate the corresponding linear estimate as closely as possible by means of sample frequencies. Since the accuracy of estimates of least squares relative weights increases with size of sample, maximum likelihood statistics have, in an asymptotic sense for large samples, the same optimum properties which are possessed in an exact sense (even for small samples) by the corresponding ideal linear estimates. Thus the results obtained by means of the theory of large samples are supported by the approach to estimation problems by means of ideal linear estimates. In addition, the later approach facilitates the integration of available techniques as explained in later sections.

It is true that the optimum properties of maximum likelihood statistics can be presented in terms of the theory of large samples, but the fact that a given method of estimation yields a statistic whose asymptotic sampling variance is a minimum does not imply that the same technique will yield a minimum variance statistic for any given small value of n. For example, it is well known that the median is a maximum likelihood estimate of the midpoint of a double exponential universe. Nevertheless, in samples of three observations from such a universe, another statistic—4/9 of the mean plus 5/9 of the median— has greater relative advantage over the median than the median has over the mean.

Fisher's discussion of the relative efficiencies of his five alternative consistent statistics suggests that it is impossible to formulate objective criteria for making choices among alternative statistics such that each statistic will be used whenever its sampling variance is smallest. Consider the sequence of universes generated by letting θ vary from zero to unity. In general, each value of θ would determine which of Fisher's five statistics would have smallest sampling variance for that particular universe for any given value of n. In comparison with any other single statistic, the statistic T_4 would usually have smaller sampling variance, but there are notable exceptions. For example, in the absence of linkage when θ is equal to one-fourth, the statistic T_2 is the ideal linear estimate and its sampling variance is smaller than that of T_4—at least for certain small values of n. For this reason, Fisher used T_2 in preference to T_4 as the basis for testing the significance of linkage. The statistic T_5—derived by Fisher's method of minimum chi-square—is also of special interest. Fisher's method of minimum chi-square yields statistics which differ from the corresponding maximum likelihood statistics because Fisher considers the denominators as variables in the process of differentiation instead of considering them as unknown parameters to be estimated by identifying them with the corresponding statistics in the numerators *after* differentiation. Arguments of later sections tend to show that the latter method is more appropriate. In this example, it can be shown that if T_5 were substituted for the corresponding parameter in the denominators of chi-square (*and treated as a parameter*) the minimization of chi-square with respect to statistics in its numerators would be exactly equivalent to substituting 0.035785, the numerical value of T_5 for θ in equation (2.5) and solving for T_6 to obtain 0.035717, a value which is much closer to 0.035712, the numerical value of the maximum likelihood estimate T_4 than to Fisher's T_5. In problems of estimation chi-square should be minimized in order to obtain efficient statistics—not to obtain a small criterion for testing goodness of fit— and it should be minimized in a manner consistent with this purpose. Whether or not it is possible to derive an even smaller value for a quantity called chi-square should be considered to be irrelevant in either estimation problems or tests of significance. It is difficult to present these ideas in more technical language because it is possible to construct trivial hypothetical universes for which Fisher's method of minimum chi-square provides statistics which are

superior in certain respects to the corresponding maximum likelihood statistics. Nevertheless, it seems clear that the ideal linear estimate usually has smaller sampling variance than the maximum likelihood statistic which, in turn, usually has smaller sampling variance than any other given practical statistic. Evidence presented in later sections tends to show that these advantages are more important in small samples than in cases in which the theory of large samples is applicable.

3. The "ideal" method of least squares. When sample observations are uncorrelated in successive samples and parameters to be estimated are linear functions of the expected values of the sample observations, the method of least squares yields ideal linear estimates of the parametes provided that the weight of each observation is inversely proportional to its variance in successive samples. Although the minimum sampling variance property among linear unbiased estimates is seldom stressed, this principle of weighting has been presented in connection with the method of least squares for more than a hundred years. In order to emphasize the theoretical nature of weights which depend on variances which are usually unknown in practice and to distinguish the method based on such weights from the more familiar method of least squares with equal weights, the method which yields ideal linear estimates will be called the *ideal method of least squares.*

Discussion of the general problem of estimating linear functions of cell proportions can be facilitated by making use of results obtained by other writers— notably Gauss (as reported by Whittaker and Robinson [6]) and Pearson [4]. According to Whittaker and Robinson, "the first writer to connect the method [of ideal least squares] with the theory of probability was Gauss" [6, p. 224]. In his Theoria Motus proof of 1809, Gauss derived the "most probable value" [6, p. 223] of a parameter (i.e., the statistic which satisfies the criterion now called maximum likelihood) for the case in which sample observations are statistically independent and normally distributed. In his Theoria Combinationis proof of 1821–23, Gauss "abandoned the 'metaphysical' basis" [6, p. 220] of his earlier work and derived the method herein called the ideal method of least squares (without approximation) from the criteria of (1) minimum variance and (2) absence of bias for the case in which "the mean value of [the covariance of a pair of errors] is zero" [6, p. 224]. Since the covariances of *uncorrelated* linear functions are zero whether they are *statistically independent* or not, it follows from the work of Gauss that the ideal method of least squares applied to uncorrelated linear functions of sample frequencies yields ideal linear estimates. In other words, the ideal method of least squares implies the following six steps:

1. From the set of $k + 1$ sample frequencies construct k linear functions which are uncorrelated in successive samples.
2. From each function subtract its expected value in terms of the unknown parameters to find its sampling error.

3. Write the ratio of each sampling error to its own standard error in the form of a fraction.
4. Sum the squares of these standardized uncorrelated sampling errors to obtain a quantity called chi-square.
5. Substitute statistics[2] for the parameters in the *numerators* of chi-square.
6. Minimize the sum of squares of residuals with respect to each statistic in turn (subject to appropriate side conditions in case linear functions not implied in preceding steps are known).

This series of six steps can be summarized by the single statement that the function to minimize is the sum of squares of standardized uncorrelated residuals. Actually this statement is oversimplified because even though sampling errors are both uncorrelated and standardized, the corresponding residuals are, in general, neither standardized nor uncorrelated.

4. Pearson's expression for chi-square. As defined by Pearson [4], chi-square is the sum of squares of a set of k standardized uncorrelated linear functions of sampling errors in a set of $k + 1$ correlated sample frequencies. A set of k standardized uncorrelated linear functions can be constructed in an infinite number of ways, but each set can be obtained from any of the others by means of an orthogonal transformation. Thus the sum of squares is the same no matter what set is originally chosen. As his set of standardized uncorrelated linear functions, Pearson chose those determined by the axes of the correlation ellipse for which he gave the required sum of squares in terms of "minors" or cofactors of the correlation determinant of the first k sample frequencies. Pearson reduced this complicated expression to the now familiar form

$$(4.1) \qquad \chi^2 = \sum_{i=1}^{k+1} (n_i - np_i)^2/np_i,$$

where p_i is the proportion in the ith cell in the universe and n_i is the frequency in the ith cell of a sample of n observations selected at random from an infinite universe (or with replacements from a finite universe).

The widespread misunderstanding of the nature of chi-square seems to be based primarily on the facts that
1. Pearson's rule for degrees of freedom is inadequate (see section 5), and
2. Pearson's expression for chi-square can be derived by approximate methods as well as by exact methods.

Pearson's derivation of the expression for chi-square by exact methods is sufficient to show that its derivation by approximate methods involves a paradox in which different sets of approximations offset each other; however, Pearson's article is relatively inaccessible and, in addition, his algabraic reductions involve

---

[2] It is convenient to call these variable symbols "statistics"; the quantities whose squares are summed, "residuals"; and the whole expression "chi-square," even though, from a certain point of view, these terms are strictly applicable only after the minimization process. This usage should always be clear from its context.

the minors of a general determinant of the kth order. For these reasons, the following exact derivation is presented in terms of elementary algebra.

Since the sum of squares is the same for any set of k standardized uncorrelated linear functions of the sampling errors in $k + 1$ correlated frequencies, a set should be chosen for which the algebraic reductions are as easy as possible. From this point of view a satisfactory set, which can be written in any of three forms, is given by

$$(4.2) \qquad y_i = p_i n_{i+} - p_{i+} n_i$$
$$= p_i e_{i+} - p_{i+} e_i$$
$$= -p_i e_{i-} - (p_i + p_{i+}) e_i$$

where $e_i = n_i - np_i$ and $i+$ and $i-$ refer to classes formed by combining all classes above the ith class and below the ith class, respectively.

By means of the known variances and covariances of the sample frequencies in expected value form,

$$(4.3) \qquad E(e_i^2) = np_i(1 - p_i),$$

and

$$(4.4) \qquad E(e_i e_j) = -np_i p_j,$$

it can be shown that the variance of y_i is

$$(4.5) \qquad E(y_i) = np_i p_{i+}(p_i + p_{i+}),$$

and, by using the third expression in equation (4.2) for y_i and the second for y_j, it can be shown that any pair of y's are uncorrelated because

$$(4.6) \qquad E(y_i y_j) = 0, \qquad\qquad (i < j).$$

Let z_i represent the variable y_i expressed in standard-deviation units. The square of this standardized uncorrelated linear function of correlated sampling errors can be written

$$(4.7) \qquad z_i^2 = \frac{(p_i e_{i+} - p_{i+} e_i)^2}{np_i p_{i+}(p_i + p_{i+})} .$$

It remains to show that Pearson's expression for chi-square can be obtained by adding the k values of z_i^2 in succession. For this purpose it is convenient to define

$$(4.8) \qquad \chi_r^2 = \sum_{i=1}^{r} \frac{e_i^2}{np_i} + \frac{e_{r+}^2}{np_{r+}} ,$$

obtained by combining all classes above the rth class.

When $r = k$, the expression in equation (4.8) is the expression to be derived. It remains to show that χ_k^2 is the sum of squares of k standardized uncorrelated linear functions of sampling errors; i.e.,

(4.9) $$\chi_k^2 = \sum_{i=1}^{k} z_i^2.$$

For the first cell $e_{1+} = -e_1$ and $p_{1+} = 1 - p_1$. Hence y_1 reduces to the negative of the error in the first frequency and

(4.10) $$\chi_1^2 = e_1^2/np_1(1 - p_1)$$

$$= e_1^2/np_1 + e_{1+}^2/np_{1+}   \qquad (p_{1+} = 1 - p_1),$$

a special case expressed in the required form. The general case is established by showing that

(4.11) $$\chi_{r-1}^2 + z_r^2 = \chi_r^2,$$

or, alternatively, that

$$z_r^2 = \chi_r^2 - \chi_{r-1}^2$$

$$= e_r^2/np_r + e_{r+}^2/np_{r+} - (e_r + e_{r+})^2/n(p_r + p_{r+})$$

(4.12)
$$= \frac{(p_{r+}e_r^2 + p_r e_{r+}^2)(p_r + p_{r+}) - p_r p_{r+}(e_r^2 + 2e_r e_{r+} + e_{r+}^2)}{np_r p_{r+}(p_r + p_{r+})}$$

$$= \frac{p_r^2 e_{r+}^2 - 2p_r p_{r+}e_r e_{r+} + p_{r+}^2 e_r^2 \equiv (p_r e_{r+} - p_{r+}e_r)^2}{np_r p_{r+}(p_r + p_{r+})},$$

thus establishing the derivation of Pearson's expression for chi-square.

When sampling is done without replacement each variance and covariance is multiplied by $(N - n)/(N - 1)$ where N is the number of observations in the universe. Hence, chi-square for this case can be written

(4.13) $$\chi^2 = \frac{N - 1}{N - n} \sum_{i=1}^{k+1} \frac{e_i^2}{np_i}.$$

This expression shows that the factor involving sampling errors is the same whether sampling is done with replacement or without replacement. Hence, the derivation of least squares statistics is the same for either method of sampling, but sampling variances for the simpler case are multiplied by the factor $(N - n)/(N - 1)$ when sampling is done without replacement.

5. The method of minimum chi-square. The derivation of Pearson's expression for chi-square completes first four steps of the ideal method of least squares outlined in section 3. Hence, the method of minimum chi-square is the sample-frequency form of the ideal method of least squares in which only two of the six steps remain to be taken.

In his original article [4] Pearson pointed out that the use of statistics instead of parameters would affect the value of chi-square but that such effects would usually be so small that no allowance need be made for them in connection with tests of significance. It is now well known that the average value of chi-square

is reduced approximately one unit for each parameter estimated from the sample, and that the main portion of this effect is on the numerators; i.e., in large samples the effect of substituting statistics for parameters in the denominators usually has a negligible effect on the value of chi-square. By confining the discussion to the case in which parameters are used in the denominators, it is possible to make simple exact statements concerning the main effects in terms of the number of squares of standardized uncorrelated linear functions—also known as the number of degrees of freedom and the mean value of chi-square.

When the expected values in the numerators of chi-square can be expressed as linear functions of r algebraically independent parameters, ideal linear estimates of the r parameters are determined by substituting statistics for the r parameters and minimizing the resulting expression wth respect to each statistic. In general, such a substitution of statistics for parameters in the numerators of chi-square reduces the number of degrees of freedom by one unit for every parameter estimated; that is, the appropriately minimized chi-square can be analyzed into $k - r$ squares of standardized uncorrelated linear functions of sampling errors.

The r ideal linear estimates are linear functions of the sample frequencies. Let $(v_1, v_2, \cdots, v_r)$ be a set of standardized uncorrelated linear functions of the correlated sampling errors in these statistics and let $(v_1, v_2, \cdots, v_k)$ be a set of linear functions obtained from the z_i's of section 3 by an orthogonal transformation. Since the sum of squares is not changed by such a transformation, chi-square is the sum of the k values of v_i^2. The process of substituting statistics for the r parameters in the numerators of chi-square reduces the values of the first rv_i^2's to zero without affecting the values of the other $(k - r)v_i^2$'s.

Thus the appropriately minimized chi-square can be analyzed into $k - r$ squares of standardized uncorrelated linear functions of sampling errors and is therefore said to have $k - r$ degrees of freedom. The mean value of each square is the variance of a standardized linear function of sampling errors and is therefore unity by definition. Hence the mean value of the appropriately minimized chi-square (with parameters in the denominators) is exactly $k - r$ when r statistics are estimated from a set of $k + 1$ sample frequencies.

The expression to be minimized is

$$(5.1) \qquad \chi^2 = \sum \frac{(n_i - m_i')^2}{np_i}$$

where m_i' is the ideal linear estimate of np_i. The set of statistics described by the equation

$$(5.2) \qquad m_i' = n_i,$$

reduces the value of chi-square to zero—its minimum value. This shows that the sample cell proportion is the ideal linear estimate of the corresponding parameter.

Whenever a linear function independent of the sum of the cell proportions is

known, it is possible to take advantage of additional information provided by the known function by minimizing chi-square subject to an appropriate side condition. When side conditions are used in this way, the number of degrees of freedom for the minimized chi-square is equal to the number of side conditions which are algebraically independent of each other (and of the sum of the cell proportions). Let the known linear function be written

$$(5.3) \qquad \Sigma a_i n p_i - m = 0.$$

In order to facilitate comparison of the typical equation of maximization with the corresponding equation of the method of maximum likelihood, it is convenient to minimize chi-square by maximizing $-\chi^2/2$ subject to a side condition based on (5.3). The function to be maximized can be written

$$(5.4) \qquad -\chi^2/2 = \Sigma(n_i - m_i')^2/(-2np_i) + h(\Sigma a_i m_i' - m),$$

where h is a Lagrange multiplier. Setting the partial derivative of $-\chi^2/2$ with respect to m_i' equal to zero, the typical equation for minimizing chi-square can be written

$$(5.5) \qquad (n_i - m_i')/np_i + ha_i = 0,$$

a form which shows that, in general, ideal linear estimates are defined in terms of unknown parameters. Fortunately, these parameters can usually be approximated closely by an iterative process. Substituting m_i for both np_i and m_i' in equations (5.5) the typical equation in the limiting values of such a process can be reduced to

$$(5.6) \qquad n_i/m_i - 1 + ha_i = 0,$$

a form which is identical with the typical equation (6.6) of maximum likelihood derived in section 6. This equality of typical equations implies that whenever the denominators of chi-square are estimated in such a way as to be consistent with least squares statistics based on them, the method of minimum chi-square always leads (by means of approximations necessary in practice) to maximum likelihood estimates of parameters which are linear functions of cell proportions.

6. The method of maximum likelihood. Maximum likelihood estimates of linear functions of cell proportions can be obtained by (1) expressing the probability function (general term of the multinomial expansion) in terms of the r parameters to be estimated; (2) substituting r statistics for the r parameters; and (3) maximizing with respect to the r statistics. In practice, this is usually accomplished by maximizing the logarithm of the variable factor in step (3) which can be written,

$$(6.1) \qquad L = \Sigma n_i \log m_i,$$

where m_i is the maximum likelihood estimate of np_i, the expected value of the ith frequency n_i in a sample of n observations classified into $(k + 1)$ classes or

cells. It is evident that L as written has no maximum with respect to any m_i since it increases without bound as m_i increases, but it sometimes has a uniquely determined maximum when each of the m_i's is written explicitly in terms of less than $k + 1$ algebraically independent statistics. In the general case it is easier to maximize L subject to an appropriate set of side conditions, one of which must be equivalent to

$$(6.2) \qquad m_1 + m_2 + \cdots + m_{k+1} - n = 0.$$

When no linear function except the sum is known, the likelihood function can be written

$$(6.3) \qquad L = \Sigma n_i \log m_i - (\Sigma m_i - n),$$

a function which, subject to equation (6.2), is always equal to that in equation (6.1) but which has a uniquely determined maximum. The typical equation of maximum likelihood, obtained by setting the partial derivative of L with respect to m_i equal to zero, is

$$(6.4) \qquad n_i/m_i - 1 = 0,$$

an equation which shows that each sample frequency is a maximum likelihood estimate of its own expected value.

When a linear function such as that in equation (5.3) is known, an improved set of maximum likelihood statistics can be found by maximizing

$$(6.5) \qquad L = \Sigma n_i \log m_i - (\Sigma m_i - n) + h(\Sigma a_i m_i - m).$$

The typical equation of maximization is found to be

$$(6.6) \qquad n_i/m_i - 1 + ha_i = 0,$$

an equation which, as stated above, is identical with equation (5.5). Since equation (5.5) was obtained as the limit of an iterative process from the typical equation (5.4) for minimizing chi-square subject to the same side condition and since each additional side condition affects the typical equation of each method in exactly the same way, the method of minimum chi-square and the method of maximum likelihood are equivalent for the general case in the sense that the method of minimum chi-square always leads to maximum likelihood statistics as limits of an iterative process.

7. Second-order tables with known expected marginal totals. As stated in section 2, the integration of available techniques is facilitated by regarding maximum likelihood statistics as the best practical approximations to the corresponding ideal linear estimates. Since this important principle may not be immediately obvious, it will be illustrated for the important special case of second-order tables for which the expected marginal totals are known.

Consider a sample of n observations arranged on two bases of classification and presented in a table containing r rows and s columns. The universe of N

observations has been completely enumerated and classified on each basis separately but not cross-classified; i.e., universe totals of first order classes are known.

For the cell in the ith row and the jth column, let p_{ij} represent the universe cell proportion; n_{ij}, the sample frequency; np_{ij}, the expected value of n_{ij}; and m_{ij}, the maximum likelihood estimate of np_{ij}. Indicating summation by substituting a dot for the letter over which summation is to be performed, the known marginal totals satisfy the equations

$$(7.1) \qquad Np_{i.} \,-\, N_{i.} \,=\, 0,$$
$$Np_{.j} \,-\, N_{.j} \,=\, 0,$$

where $p_{i.}$ and $p_{.j}$ are the universe proportions and $N_{i.}$ and $N_{.j}$ are the known universe totals in the ith row and the jth column, respectively.

When n observations of a random sample are arranged according to two bases of classification in a table with r rows and s columns for which the $r + s$ marginal totals are known, the typical equation of maximum likelihood can be obtained by maximizing, subject to side conditions based on equations (7.1), the likelihood function

$$(7.2) \qquad L \,=\, \Sigma\Sigma n_{ij} \log m_{ij} \,-\, \Sigma a_i(m_{i.} \,-\, n_{i.}) \,-\, \Sigma b_j(m_{.j} \,-\, n_{.j}),$$

with respect to the maximum likelihood estimates m_{ij}, where a_i and b_j are typical Lagrange multipliers. Setting the partial derivative with respect to m_{ij} equal to zero and transposing, the typical equation of maximum likelihood can be written

$$(7.3) \qquad n_{ij}/m_{ij} \,=\, a_i \,+\, b_j \,.$$

Since equations (7.3) are not linear in their unknowns, the reader's first reaction might well be to agree with a certain anonymous critic that "their solution is difficult." This impression of great difficulty is probably the chief reason that previous writers have not used the method of maximum likelihood for this type of problem even after they had developed a set of techniques adequate for the solution of the equations of maximum likelihood. In other words, all that was needed was the integration of available techniques as will now be shown.

In 1940, Deming and Stephan [2] derived a set of normal equations for the adjustment of a set of second-order cell frequencies to known expected marginal totals by the method of least squares in which each sample frequency is weighted by its own reciprocal. This method yields statistics which are efficient according to the theory of large samples, but they do not satisfy the criterion of maximum likelihood exactly. In the same article was presented an easier method of *iterative proportions*, which, unfortunately, does not yield least squares statistics. In 1942, Stephan [5] developed an improved iterative process which yields statistics which satisfy the criterion of least squares with arbitrarily

chosen weights. The foregoing developments are presented in greater detail in Deming's book [1] in which Deming adapts Stephan's iterative method to the particular case in which each sample frequency is weighted by its own reciprocal so as to yield solutions for the normal equations derived in the joint article [2].

In Deming's notation, equation 8 of Stephan's article [5, p. 169] can be written

$$(7.4) \qquad m_{ij} = c_{ij}(p_i + q_j - 1) + n_{ij},$$

an expression obtained by substituting c_{ij} for np_{ij} in the denominators of chi-square and minimizing with respect to the statistics in the numerators. Hence, if exact values of the np_{ij} were used for the c_{ij}, the Stephan iterative method would yield ideal linear estimates. Unless these parameters are implied by some hypothesis to be tested, it is necessary, in practice, to estimate the np_{ij} from sample data. In order to secure maximum likelihood estimates of expected cell frequencies by means of the Stephan iterative method, the adjusted frequencies based on first approximations to the c_{ij} should be used as second approximations to the c_{ij}, etc. In this way, maximum likelihood statistics can be derived to any desired degree of approximation. At this point it should be emphasized that the preceding statement applies not only to the class of problems considered in this section but also to the wider class of problems for which the Stephan iterative method provides solutions.

Unfortunately, theoretical discussions of previous writers contain confusing compensating errors which (1) present their own methods in an unnecessarily unfavorable light and (2) increase the difficulties involved in the introduction of the improvements in techniques suggested in section 9 which involve some degree of adaptation of techniques already available. For these reasons, it seems necessary to follow the arguments of previous writers in order to show the points at which improvements are needed. This can be done most effectively in connection with Deming's book [1] where the method of least squares is presented in great detail.

For the special case in which the sampling errors in the observations are uncorrelated, the ideal criterion of least squares implies that the weight of each observation should be inversely proportional to its sampling variance. This criterion is accepted as well known by Deming who says that "the principle of least squares requires the minimizing of the sum of the weighted squares of the residuals" [1, p. 14] where "the weights of two functions are inversely proportional to their variances" [1, p. 22]. Deming assumes that "there is no correlation between the errors in the observations" with the qualification that "this assumption covers a wide class of problems, but does fail to cover some." [1, p. 49]. This assumption of uncorrelated errors is not applicable to sample-frequency problems, of course, because the sample frequencies are correlated with each other in such a way that the reciprocals of the ideal least squares weights are not proportional to the sampling variances $np_{ij}q_{ij}$ but rather to the expected frequencies np_{ij} which appear in the denominators of chi-square.

In this connection it is interesting to note that Deming himself insists that "there is only one principle of least squares, namely, the minimizing of χ^2." [1, p. 51]. However, the method currently in use for the minimizing of chi-square was that given by Fisher [3] which leads to equations which are difficult to solve even for such a simple example as the one presented in section 2 above.

Deming and Stephan are to be commended for seeking an easier method but there is no justification (even as a device for saving effort) for their modification of the "principle of least squares" so as to imply erroneously that

(1) weights of correlated sample frequencies are inversely proportional to their variances, and

(2) sample frequencies are, in general, approximately proportional to their own sampling variances.

Strangely enough, these two errors were applied in combination by Deming and Stephan to obtain good practical approximations to the ideal least squares weights. It might be argued that the second misleading implication is really not an error because it is offered as a simplifying approximation, but it is an integral part of both the normal equations approach in the joint article [2] and Deming's adaptation [1] of the Stephan iterative method; that is, in each case the method would have to be revised if better approximations to the ideal least squares weights were used. More explicitly, Deming (1) uses n_{ij} for Stephan's c_{ij} in equation (7.4); (2) identifies it with the other n_{ij} in the same equation; and (3) reduces the equation to a different form thus effectively preventing the use of successive approximations to the c_{ij} without returning to Stephan's iterative method in the general form given by equation (7.4) above which Deming does not present at all. Results of the joint article [2] are quoted by Stephan [5] without any explanation of the nature of the errors, but none of these results are used in the development of his iterative method which as noted above, is applicable to any arbitrarily chosen set of weights. The fact that Stephan corrected the second error without correcting the first implies that the weights he actually used are unsatisfactory. In Deming's adaptation of the Stephan iterative method, a much better set of weights is obtained, not by correcting the first offsetting error overlooked by Stephan, but by resurrecting the second offsetting error which Stephan had corrected. Since this error is an integral part of Deming's adaptation, Deming's theoretical discussion implies that his own efficient statistics are only rough approximations which are definitely inferior to the inefficient statistics obtained by means of the weights chosen by Stephan. These inconsistencies are most clearly brought out by Deming when he says:

"Strictly, in random sampling, the reciprocal of the weight of n_{ij} is $np_{ij}q_{ij}$, which is nearly equal to $n_{ij}q_{ij}$ where p and q have their usual connotations. But since factors proportional to the weights may be substituted for them, it is sufficient to use n_{ij} as the reciprocal of the weight in cell ij, since the values of q_{ij} do not usually vary much over the table." [1, p. 102.]

In any given problem the seriousness of the error in the first statement in the foregoing quotation depends on the variation among the q_{ij}'s. In the par-

ticular example used by Deming the error is of considerable importance because the largest q_{ij} is more than 40 per cent larger than the smallest q_{ij}. The weights actually used by Deming agree with weights implied by the ideal method of least squares except for sampling errors in the n_{ij}; hence, the error in any relative weight converges stochastically to zero so that Deming's statistics are efficient according to the theory of large samples. The efficiency of Deming's statistics is inconsistent with the theory presented by Deming which implies erroneously that efficiency of estimation depends on approximate equality of cell proportions. If this argument were true it would apply also to the method of maximum likelihood and all other methods which yield efficient practical statistics in sample-frequency problems. The foregoing discussion, together with the results of section 8 show that the theory as presented by Deming has the following seriously misleading features:

(1) it is based on a paradox in which a good final result is obtained by means of compensating errors;

(2) it presents his efficient statistics in an unnecessarily unfavorable light;

(3) it emphasizes the irrelevant condition of approximate equality of universe cell proportions;

(4) it fails to mention the important condition of proportionality by rows and columns; and

(5) it makes least squares, minimum chi-square, and maximum likelihood seem to be competing alternative methods.

Of these undesirable characteristics, the last two are probably the most serious because they make the effective integration and adaptation of statistical techniques more difficult. As has been shown in sections 4, 5, and 6, the sample-frequency form of the ideal method of least squares is the method of minimum chi-square which always leads (by means of appropriate practical approximations to unknown weights) to maximum likelihood statistics; in other words, the methods are equivalent from a practical point of view.

Since the ideal method of least squares based on the unknown np_{ij} determines fully efficient, but theoretical, ideal linear estimates, the efficiency of practical approximations to ideal linear estimates depends on the accuracy with which the denominators of chi-square are estimated. For the unknown denominators np_{ij}, Deming uses the sample frequencies n_{ij} while the method of maximum likelihood implies the use of the corresponding maximum likelihood estimates— statistics which, in general, have smaller sampling variances. The foregoing argument suggests that maximum likelihood statistics are slightly superior to Deming's statistics for any given finite value of n and that their relative advantage increases as the sample size decreases. In large samples both methods yield efficient statistics because the relative errors in the weights implied by either method converge stochastically to zero as n increases. Although the advantage of maximum likelihood statistics over Deming's statistics is unimportant except in small samples, it can be shown that Deming's choice of weights leads to imperfectly compensated negative errors of estimation even in his large sample of 33,837 observations.

Deming weights each sample frequency by its own reciprocal. Positive errors of sampling decrease the value of the reciprocal and thus increase the absolute size of the required negative adjustments. Negative errors of sampling increase the value of the reciprocal and thus decrease the size of the positive adjustment. Thus every error of sampling (either positive or negative) leads to a negative error of estimation due to inappropriate weighting. Because the sum of all adjustments must be zero, these negative errors of estimation are compensated on the average but more or less imperfectly. The net effect of this imperfect compensation of negative errors of estimation is that Deming's statistics are too small in those cells in which the relative adjustments (either positive or negative) are large, and vice versa. In a preliminary draft of this article, this type of error of estimation was studied by comparing Deming's statistics with the corresponding maximum likelihood statistics in conection with Deming's example involving 33,837 observations. Although errors of estimation of the type under discussion are apparent, they are, of course, extremely small in such a large sample. For this reason the large-sample comparson has been deleted in favor of simple hypothetical examples designed to throw light on similar errors of estimation in statistics derived by Fisher's method of minimum chi-square as well as in those derived by Deming's adaptation of Stephan's iterative method.

Consider a set of sample frequencies in a two-by-two table for which all expected marginal totals are equal. For this special case, the cell proportions on each diagonal are equal and the ideal linear estimate (which is also the maximum likelihood estimate) of any cell proportion is the mean of the two sample cell proportions on its diagonal. For the same case, Deming's adaptation of the Stephan iterative method yields an estimate for each cell which is proportional to the harmonic mean of sample proportions on its diagonal while Fisher's method of minimum chi-square yields estimates proportional to the corresponding quadratic means.

As a numerical example of the foregoing problem consider the set of frequencies

$$(7.5) \qquad\qquad n_{11}, n_{12}, n_{21}, n_{22} = 1, 4, 3, 2,$$

obtained in a sample of 10 observations selected at random from a universe in which the cell poportions are known to be

$$(7.6) \qquad\qquad p_{11}, p_{12}, p_{21}, p_{22} = p, 0.5 - p, 0.5 - p, p.$$

As estimates of the parameter p, the ideal linear estimate is .15, Deming's adaptation of the Stephan iterative method yields .14, and Fisher's method of minimum chi-square yields .1545 to four decimal places, the other two estimates being exact. The results illustrate the imperfectly compensated errors of estimation explained previously. The two sample frequencies on the principal diagonal (n_{11} and n_{22}) have greater relative dispersion than the frequencies on

the other diagonal. For this reason, the relative adjustments made by Deming's method are greater and according to the principle of imperfectly compensated negative errors of estimation, the estimate of p obtained by Deming's method is smaller than the ideal linear estimate of p. Fisher's method of minimum chi-square yields an estimate of p which is greater than the ideal linear estimate. In fact, one should usually expect imperfectly compensated errors of estimation in statistics derived by Fisher's method of minimum chi-square to be opposite in sign and about half as large as those in the corresponding statistics derived by means of Deming's adaptation of the Stephan iterative method.

At this point, it should be emphasized that Fisher does not recommend his own method of minimum chi-square in preference to the method of maximum likelihood. In fact, he presents the theory of estimation in such a way as to imply correctly that the method of maximum likelihood is superior, especially in small samples. Other writers have noted the small differences between equations of maximum likelihood and those for minimizing chi-square by Fisher's method and some have even derived one set of equations from the other by neglecting higher order terms in a Taylor series expansion. These derivations are of no interest here because they seem to justify the method of maximum likelihood as a simple approximation to some more complicated method. This type of justification is both unnecessary and undesirable. It is more useful to regard the method of maximum likelihood as an approximation to a method— least squares—for which the theory is simpler.

Skeptical readers who find the foregoing argument unconvincing may be able to profit from the following example. Consider the problem of estimating the parameter p where $2p$ is the proportion of white balls in an urn. A sample of 10 balls is selected and classified by the following process. Each white ball is placed in one of the cells on the principal diagonal of a two-by-two table, the particular cell being decided by the toss of a coin. A similar method is used for non-white balls placed in cells on the other diagonal. Assuming that the results of this process are given by equation (7.5), which of the three alternative estimates of p given above should be preferred? Belief in the general superiority of Fisher's method of minimum chi-square seems to imply that the device of coin-tossing described in this example can be used in practical problems involving the estimation of the proportion of "successes" to secure estimates which are superior to the sample proportion—the ideal linear estimate in such cases. Even if it is possible to construct trivial special case examples supporting some complicated method for such problems the general use in practical problems of the coin-tossing device in connection with either Fisher's method of minimum chi-square or Deming's adaptation of the Stephan iterative method would be absurd as this example is intended to emphasize.

8. The method of proportional distribution of marginal adjustments. The method of proportional distribution of marginal adjustments is a general method of adjusting sample frequencies so that their row and column totals agree with

known expected marginal totals. In other words, the adjusted frequency for the cell in the ith row and the jth column is given by the equation

(8.1) $$m_{ij}^{*} = n_{ij} - p_{i.}d_{.j} - p_{.j}\,d_{i.}\,,$$

where

$$d_{i.} = m_{i.} - n_{i.}\,,$$

and

$$d_{.j} = m_{.j} - n_{.j}\,,$$

are the net adjustments in the sample cell frequencies of the ith row and the jth column, respectively. The asterisk is used to distinguish maximum likelihood estimates m_{ij} and the ideal linear estimates m_{ij}' from the set of statistics based on equation (8.1).

The method of proportional distribution of marginal adjustments yields ideal linear estimates when the universe cell proportions are proportional by rows and by columns; i.e., when

(8.2) $$p_{ij} = p_{i.}p_{.j}\,.$$

This important principle can be established by substituting in equation (7.4) of section 7 the quantities

(8.3) $$c_{ij} = np_{i.}p_{.j}\,,$$

$$p_i = 0.5 + d_{i.}/np_{i.}\,,$$

and

$$q_j = 0.5 + d_{.j}/np_{.j}\,,$$

and reducing the typical equation of the ideal method of minimum chi-square to the form of equation (8.1) which defines the method of proportional distribution of marginal adjustments.

Even in the absence of exact proportionality, under which it yields fully efficient statistics, the method of proportional distribution of marginal adjustments has the following relative advantages over other available methods:

(1) ease of extension to tables of higher order;

(2) *exact* agreement with known (expected) marginal totals;

(3) simplicity of interpretation;

(4) independence of computational errors;

(5) rapidity of processing;

(6) economy of effort; and

(7) fully efficient criteria for testing the significance of departures from proportionality of rows and columns.

Ease of extension to tables of higher order is a desirable property of the method of proportional distribution of marginal adjustments. Equation (8.1)

applies to the special case in which there are only two bases of classification. In the more general case sample observations are cross-classified according to r bases of classification, each cell frequency in an rth order table being the number of observations in the corresponding rth order class whose expected value is to be estimated. The required adjustment for each first order class (obtained by subtracting the sample total from its known expected value) is distributed among the various cells in proportion to the universe totals of the corresponding $(r - 1)$th order classes to which the cells belong. The general process is illustrated by

$$(8.4) \qquad m^*_{ijk} = n_{ijk} + p_{i..}d_{.jk} + p_{.j.}d_{i.k} + p_{..k}d_{ij.} ,$$

the formula for estimating the expected frequency in the general cell of a third order table.

Exact agreement with marginal totals follows easily from the method of proportional distribution and can be established algebraically by summing the estimation equation by first order classes; e.g., summing equation (8.1) by rows and columns. In practice, discrepancies are always either errors of rounding or mistakes in computation; they are never due to lack of convergence of iterative processes as is often true in alternative methods of estimation.

Although simplicity of interpretation is desirable in general, it is especially important when random sampling is an unrealistic abstraction. For example, the method of proportional distribution of marginal adjustments has been used to estimate the cell proportions in a two-way classification of incomes from known marginal proportions and a detailed cross classification at an earlier date. In this problem known shifts in income distributions made it evident that certain cells previously vacant should not have the zero proportions which would be estimated for them by other available methods of estimation. The ease with which the effects of the method of adjustment can be traced is important also in the analysis of the results of sample surveys in which various types of bias are important.

The method of proportional distribution of marginal adjustments yields the estimated expected frequency for any cell by a single sequence of computations which is independent of the corresponding process for any other cell. Errors made in computing the estimate for any cell appear in marginal totals of estimates for all first order classes which include that cell. If only a few errors are made in a table they can be localized immediately and can be corrected without recomputing any estimates which are correct.

In certain types of social surveys, rapidity of processing is so important that, as Deming puts it, "the delay of only the brief time required for adjustment may not be advisable." [1, p. 102]. Under these conditions, it is important to have a simple formula like equation (8.1) in which substitutions can be made rapidly. Even when the time element is relatively unimportant, the economy of effort and the ease of explaining the method to clerical assistants are often of practical importance.

Finally, departures from proportionality among rows and columns often provide the chief element of interest in research studies—not only in social surveys of the type illustrated in Deming and Stephan's example but also in biological sciences. The most effective tests of significance for the purpose of presenting statistical evidence of lack of proportionality are those based on statistics like those derived by the method of proportional distribution of marginal adjustments whose efficiency is 100 per cent when proportionality is exact.

Even when proportionality is not exact, the efficiency of statistics derived by proportional distribution may be close to 100 per cent under fairly typical problem conditions such as those in the example by Deming and Stephan wherein the other more complicated methods require several times as much computational effort, but have little advantage over the easier method with respect to efficiency of estimation in this particular problem.

9. Suggested improvements in techniques. In section 7, a method was outlined by which it is possible to derive sets of maximum likelihood statistics by merely integrating available techniques without changing any of them. In this section a number of improvements are suggested. At this point it should be emphasized that a given change is not an improvement merely because it yields slightly more accurate estimates or makes possible a slight saving of time and effort. In each case the research worker should consider saving of time and effort and accuracy of estimation simultaneously. In particular, it seems likely that most social surveys of the type considered by Deming and Stephan are characterized by approximate proportionality by rows and by columns— conditions relatively favorable to the simple method of proportional distribution of marginal adjustments. It should be clearly understood that suggestions in this section are intended for those research workers whose problems justify a great deal more effort than is required to adjust sample frequencies by this simple method.

Assuming that the problem at hand warrants the effort required to derive maximum likelihood estimates, the first consideration is the derivation of a set of $m_{ij}(1)$, first approximations to the m_{ij}, and a set of values of $p_i(1)$, first approximations to the p_i. Even if proportionality by rows and by columns is not closely approximated use of values of the $p_i(1)$ provided by equation (8.3) are especially to be recommended. In the example used by Deming these values for the $p_i(1)$ are so much better than the values recommended by Deming that they save a large proportion of the effort required by the iterative process. If rows and columns are approximately proportional, equation (8.1) should be used to provide values of the $m_{ij}(1)$, in which case it is possible to use an iterative process similar to the one used by Deming but based on the typical equation of maximum likelihood (7.3) to achieve a given degree of accuracy in the maximum likelihood estimates with even less effort. Under favorable conditions such as those in Deming's example the suggested iterative process yields excellent

approximations to maximum likelihood estimates by means of the following steps:

1. Construct a set of first approximations to the r row components of the rs maximum likelihood divisors $(a_i + b_j)$ by means of the equation

$$(9.1) \qquad a_i(1) = n_{i.}/np_{i.} - 1/2.$$

2. Compute successive approximations to the a_i and b_j by means of the equations

$$(9.2) \qquad b_j(g) = [n_{.j} - \Sigma m_{ij}(1)a_i(g)]/np_{.j},$$

$$(9.3) \qquad a_i(g + 1) = [n_{i.} - \Sigma m_{ij}(1)b_j(g)]/np_{i.},$$

where $m_{ij}(1)$, the first approximation to m_{ij}, is derived by means of equation (8.1). Just as in Deming's iterative process, the expression in brackets is a series of products which can be subtracted in a single sequence of machine operations and the final division can be performed without having to record any of the intermediate results.

3. Divide the sample frequencies by the maximum likelihood divisors to obtain the maximum likelihood estimates

$$(9.4) \qquad m_{ij} = n_{ij}/(a_i + b_j),$$

where limiting values of a_i and b_j are approximated as closely as desired by successive approximations in the preceding equations.

Under unfavorable conditions, the iterative process of this section is not always the easiest way to obtain satisfactory estimates. For example, when samples are small and/or rows and columns are not approximately proportional, it is better to use the iterative method as originally presented by Stephan where sample frequencies can be used for first approximations to the c_{ij} and these may be replaced by successively better approximations.

The point made in the final paragraph of Fisher's well-known book [3] that "in practice one need seldom do more than solve, at least to a good approximation, the equation of maximum likelihood," is strongly supported by the developments of this article. In addition, the proof that the method of least squares and the method of minimum chi-square always lead (by means of approximations to ideal weights) to maximum likelihood statistics greatly facilitates the adaptation of techniques developed in connection with these hitherto competing methods.

REFERENCES

[1] W. EDWARDS DEMING, *Statistical Adjustment of Data*, John Wiley & Sons, 1943.

[2] W. EDWARDS DEMING AND FREDERICK F. STEPHAN, "On a least squares adjustment of a sample frequency table when the expected marginal totals are known," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 427–444.

[3] R. A. FISHER, *Statistical Methods for Research Workers*, 6th ed., Oliver and Boyd, 1936, Ch. 9.

[4] KARL PEARSON, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Phil. Mag.*, Vol. 50 (1900), pp. 157–175.

[5] FREDERICK F. STEPHAN, "An iterative method of adjusting sample frequency tables when expected marginal totals are known," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 166–178.

[6] E. T. WHITTAKER, AND G. ROBINSON, *The Calculus of Observations*, D. Van Nostrand Company, 1924, Ch. 9.

A STATISTICAL PROBLEM CONNECTED WITH THE COUNTING OF RADIOACTIVE PARTICLES

By Sten Malmquist

Institute of Statistics, University of Upsala, Sweden

1. Introduction. Our problem refers to random events forming a sequence in time or in space, *e.g.* particles emitted by a radioactive matter. By omitting certain elements of the given sequence, say f, we form another sequence, say g. The rule of omission involves an arbitrarily prescribed constant u. The rule to be followed in forming g is:

Case I: Let a be an element in f and g. The next element to be included in g is then the first element in f which follows a after a distance greater than u.

Case II: Let a be an element in f and g. The next element to be included in g is then the first element in f which follows a at a distance greater than u from the preceding element in f, whether this belongs to g or not.

When the events are represented by impulses emitted by a radioactive matter and feeding a recorder with a constant resolving time u, the new sequence consists of the counted impulses. The two cases correspond to the reaction of different types of recorders. The distinction between the two transformations has caused some confusion. It has, however, been clearly pointed out by Ruark and Brammer [5].

v. Bortkiewicz [2] seems to be the first who has considered problems related to the transformed sequence. Starting from investigations by Rutherford, Geiger, and others, concerning the number of recorded α-particles during a certain interval of time, say T, he observed that the distribution of this number was similar to that of Poisson but with a slightly smaller dispersion. This fact he supposed to be caused by a constant resolving time u of the recorder. By means of certain assumptions he tried to calculate the effect on the mean and the dispersion by the transformation in Case I, supposing the cumulative distribution function $F(t)$ for the distance between two consecutive elements in the sequence f is given by

$$F(t) = 1 - e^{-at},$$

where here and in what follows, t denotes a non-negative variable.

Considering Case II with $F(t)$ as above, Levert and Scheen [4] have recently worked out an expression for the distribution of the number of elements during T in the sequence g.

Gnedenko [3] has considered the distribution of the number of lost elements in Case I with particular regard to the initial state of rest.

Alaoglu and Smith [1] considered problems referring to successive transformations of a sequence. When, for example, a sequence of particles enters a tube-counter and amplifier, together acting with a resolving time u_1, and

the impulses then are feeding a recorder with resolving time $u_2 > u_1$, the sequence of recorded impulses will be the result of two successive transformations. If we have a scaling circuit between the counter and the recorder, we have to make a transformation of another type between the two transformations in Case I and Case II.

The present paper deals with the transformed sequence in Case I. The distribution function $F(t)$ is supposed to be arbitrary. An advantage of this generalization is that the formulas derived could be used in treating problems referring to successive transformations.

The author wishes to express his sincere gratitude to Professor Herman Wold for stimulating discussions and valuable advice.

2. Derivation of distributions for case I. Suppose that the sequence f has $F(t)$ for distribution function for the distance between two consecutive elements. $F(t)$ is supposed to be independent of absolute time (space), and of the preceding distance between two elements. When not stated otherwise, we further suppose $F(0) = 0$.

Now let $G(t)$ be the distribution function for the distance between two consecutive elements in the transformed sequence g. Evidently $G(t)$ also is independent of absolute time and of the preceding distance between two elements.

We shall consider certain distribution functions connected with $F(t)$. These functions will then be used in solving problems concerning the sequence g.

Let $F_n(t)$ be the distribution function for the distance between the first and the last of $n + 1$ consecutive elements in the sequence f. Then $F_n(t)$ is given by the recursive system

(1)
$$F_{m+n}(t) = \int_0^t F_m(t - x) \, dF_n(x); \qquad (m, n \geq 1)$$

$$F_1(t) \equiv F(t).$$

As is easily seen, we have

$$F_{m+n}(t) \leq F_m(t) \cdot F_n(t);$$

and, for $t = u$,

$$F_n(u) \to 0, \qquad \text{as } n \to \infty;$$

$$\sum_{n=1}^{\infty} F_n(u) < \infty, \qquad \text{provided that } F_1(0) < 1.$$

Alternatively, $F_n(t)$ could be deduced by the use of characteristic functions.

Still considering the sequence f, let $\Phi(t)$ be the distribution function for the distance d between an arbitrarily chosen point and the following element. Suppose that the arbitrary point is chosen so that the distance between the pre-

ceding and the following element is x. Under this condition we have, in usual symbols,

$$P(d > t) = \frac{x - t}{x}.$$

Hence,

$$\Phi(t) = 1 - \int_t^\infty \frac{x - t}{x} \, dH(x)$$

where $H(t)$ is the distribution function for the distance x.

To deduce $H(t)$ we suppose that the distribution $F(t)$ has a finite mean,

$$m = \int_0^\infty t \, dF(t).$$

By the definition of $H(t)$, we then have

$$H(x) = \frac{1}{m} \int_0^x t dF(t).$$

Thus

(2) $$\Phi(t) = \frac{1}{m} \left[\int_0^t x \, dF(x) + t \int_t^\infty dF(x) \right].$$

The corresponding frequency function $\varphi(t)$ is given by

$$\varphi(t) = \frac{1 - F(t)}{m}.$$

Consider $n + 2$ consecutive elements in f, say $a_0, a_1, \cdots, a_{n+1}$, where a_0 is an element in the transformed sequence g. The probability P_n that the next element in g following a_0 will be a_{n+1} is given by

$$P_n = F_n(u) - F_{n+1}(u), \qquad (n = 1, 2, \cdots),$$

$$P_0 = 1 - F(u).$$

Now let $P_n(t)$ be the probability that the distance between a_0 and a_{n+1} is smaller than or equal to t, when a_0 an a_{n+1} are two consecutive elements in the sequence g. Then

$$P_n(t) = \frac{1}{F_n(u) - F_{n+1}(u)} \int_0^u \left[F(t - x) - F(u - x) \right] dF_n(x),$$

$$(n = 1, 2 \cdots), \qquad P_0(t) = \frac{F(t) - F(u)}{1 - F(u)}.$$

Let $G^*(t)$ be defined by

$$G^*(t) = \sum_{n=0}^\infty P_n \cdot P_n(t) = F(t) - F(u)$$

$$+ \sum_{n=1}^\infty \int_0^u [F(t - x) - F(u - x)] \, dF_n(x); \qquad t > u.$$

When $G^*(t)$ is a distribution function, then $G^*(t)$ equals $G(t)$.

For $t_1 < t_2$ we obviously have $G^*(t_1) \leq G^*(t_2)$.

For $t = \infty$

$$G^*(\infty) = 1 - F(u) + \sum_{n=1}^{\infty} \int_0^u [1 - F(u - x)] \, dF_n(x)$$

$$= 1 - F(u) + \sum_1^{\infty} F_n(u) - \sum_1^{\infty} F_{n+1}(u) = 1.$$

Hence we take

(4)
$$G(t) = G^*(t); \qquad\qquad t > u,$$
$$G(t) = 0; \qquad\qquad t \leq u.$$

When the corresponding frequency functions $g(t)$ and $f(t)$ exist, we get

(5)
$$g(t) = f(t) + \sum_{n=1}^{\infty} \int_0^u f(t - x) f_n(x) \, dx; \qquad\qquad t > u.$$

Dealing with a sequence of elements we are often concerned with the number of occurrences during a certain time T.

Let the mean number of occurrences during T be $M(T)$. Supposing that the mean $m = \int_0^{\infty} t \, dF(t)$ is finite and that $F(0) < 1$, we have

(6)
$$M(T) = T/m.$$

We define

$$K_1(t) = \begin{array}{ll} F(t) & \text{for } t \geq \epsilon \\ 0 & \text{for } t < \epsilon \end{array}$$

$$K_2(t) = \begin{array}{ll} F(t) & \text{for } t \geq \epsilon \\ F(\epsilon) & \text{for } t < \epsilon \end{array}$$

and denote the corresponding means by $M_1(T)$ and $M_2(T)$. As is easily seen,

$$M_1(\epsilon) \leq M(\epsilon) \leq M_2(\epsilon).$$

Using (2),

$$M_1(\epsilon) = \frac{\epsilon F(\epsilon) + \epsilon[1 - F(\epsilon)]}{\int_0^{\infty} x \, dK_1(x)} = \frac{\epsilon}{\int_0^{\infty} x \, dK_1(x)} \; ;$$

$$M_2(\epsilon) = \frac{1}{\int_0^{\infty} x \, dK_2(x)} \cdot [1 \cdot \epsilon[1 - F(\epsilon)]^2 + \cdots + n \cdot \epsilon[1 - F(\epsilon)]^2 F(\epsilon)^{n-1} + \cdots]$$

$$= \frac{\epsilon}{\int_0^{\infty} x \, dK_2(x)} \cdot$$

Making $N = T/\epsilon$ and summing, we obtain

$$M_1(T) = \frac{T}{\int_0^\infty x\, dK_1(x)} = \frac{T}{m - \int_0^\epsilon x\, dF(x) + \epsilon F(\epsilon)};$$

$$M_2(T) = \frac{T}{\int_0^\infty x\, dK_2(x)} = \frac{T}{m - \int_0^\epsilon x\, dF(x)}.$$

By choosing ϵ arbitrarily small, we get

$$M(T) \to T/m.$$

Let $P(n, T)$ be the probability that we get n elements in f during a time T. Suppose that the first of these elements, a_1, comes at $T_0 + x$, and the last, a_n, at $T_0 + x + y$.

We then have

(7) $$P(n, T) = \int_0^T \varphi(x)\, dx \int_0^{T-x} [1 - F(T - x - y)]\, dF_{n-1}(y).$$

In (4) and (7) we have equations for the transformation in Case I. Because of the general form of $F(t)$, the formulas also can be used when we are concerned with successive transformations. It can further be remarked that the transformation of a sequence of impulses by passing a scaling circuit is expressed by the system (1).

3. Results for a particular form for $F(t)$. The preceding formulas will now be used for a special distribution function $F(t)$. Suppose that the frequency function $f(t) = dF(t)/dt$ is equal to the frequency function of the distance between an arbitrary point and the following element.

From (3) we get

$$F'(t) = \frac{1 - F(t)}{m},$$

or, when $F(0) = 0$,

(8) $$F(t) = 1 - e^{-at};$$

(9) $$f(t) = ae^{-at}, \quad \text{where } 1/a = m = \int_0^\infty tf(t)\, dt.$$

By means of the theory of characteristic functions we have

(10) $$f_n(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\eta(x)]^n\, e^{-itx}\, dx; \qquad\qquad f_1(t) = f(t);$$

where

(11) $$\eta(x) = a \int_0^\infty e^{-at}\, e^{itx}\, dt = \frac{a}{a - ix}.$$

Thus

$$(12) \qquad f_n(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{a^{\cdot\cdot}}{(a - ix)^n} e^{-\imath t x} dx$$

For $n = 1$, we get

$$(13) \qquad f_1(t) = ae^{-at} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{a}{a - ix} e^{-\imath t x} dx$$

By differentiating (13) $n - 1$ times with respect to a we obtain

$$(-t)^{n-1} e^{-at} = \frac{1}{2\pi} (-1)^{n-1}(n - 1)! \int_{-\infty}^{+\infty} \frac{e^{-\imath t x}}{(a - ix)^n} dx.$$

Hence, from (12),

$$(14) \qquad f_n(t) = \frac{a^n}{(n - 1)!} t^{n-1} e^{-at}.$$

From (5) we obtain the frequency function for the transformed sequence g

$$(15) \qquad g(t) = ae^{-at} + \sum_{n=1}^{\infty} \int_0^u ae^{-at} \frac{a^n}{(n - 1)!} t^{n-1} dx = ae^{au} e^{-at}; \qquad t \geq u$$

$$G(t) = 0; \ t < u.$$

The mean m_g is given by

$$m_g = a \int_u^{\infty} te^{au} e^{-at} dt = \frac{1}{a} + u.$$

Remark: Suppose the constant u is allowed to vary independently of t and that the frequency function of u is $\gamma(u)$, we obtain

$$m_g = \int_0^{\infty} t \, dt \int_0^t g(u, t)\gamma(u) \, du = \int_0^{\infty} \frac{1}{a} \gamma(u) \, du + \int_0^{\infty} u\gamma(u) \, du$$

$$(16)$$

$$= \frac{1}{a} + m(u).$$

Now let the sequence of elements, g, by means of (5) be transformed into a new sequence, h. When we are concerned with the counting of particles, emitted from a radioactive matter, let the sequence g consist of impulses from a counter-amplifier with resolving time u, feeding a recorder with resolving time u_1. Then the elements in h are the counted impulses, it being supposed that the tube-counter and the recorder reacts according to the assumptions.

We suppose $u_1 > u$. When $u_1 \leq u$, the sequences g and h are identical.

Let $g_n(t)$ denote the frequency function of the distance between the first and the last of $n + 1$ consecutive elements in g. We find, in the same way as used in obtaining (14),

$$(17) \qquad g_n(t) = \frac{a^n}{(n - 1)!} e^{anu}(t - nu)^{n-1} e^{-at}; \qquad t \geq nu.$$

Let $h(t)$ be the frequency function for the distance between two consecutive elements in the sequence h. Let further N be the greatest integer smaller than or equal to u_1/u.

Using (4) and (5) we obtain

$$h_I(t) = ae^{au} e^{-at} \sum_0^N \frac{a^n}{n!} (u_1 - nu)^n e^{anu}; \qquad t \geq u_1 + u;$$

$$(18) \quad h_{II}(t) = ae^{au} e^{-at} \sum_0^N \frac{a^n}{n!} [t - (n+1)u]^n e^{anu}, \qquad (N+1)u \leq t \leq u_1 + u;$$

$$h_{III}(t) = ae^{au} e^{-at} \sum_0^{N-1} \frac{a^n}{n!} [t - (n+1)u]^n e^{anu}, \qquad u_1 \leq t \leq (N+1)u.$$

The mean m_h is found to be

$$(19) \qquad m_h = \left[\frac{1}{a} + u\right]\left[1 + \sum_{n=1}^N \sum_{v=n}^\infty \frac{(u_1 - nu)^v a^v}{v!} e^{-a(u_1-nu)}\right].$$

We also have

$$\int_{u_1+u}^\infty th_I(t)\, dt < m_h < \int_u^\infty th_I(t)\, dt$$

or

$$\left[\frac{1}{a} + u_1 + u\right]\left[\sum_0^N \frac{a^n}{n!} (u_1 - nu)^n e^{-a(u_1-nu)}\right]$$
$$< m_h < \left[\frac{1}{a} + u_1\right] e^{au} \left[\sum_0^N \frac{a^n}{n!} (u_1 - nu)^n e^{-a(u_1-nu)}\right].$$

We now consider the number of occurrences during a time interval T. Using (6), (16), and (19) we immediately get the mean numbers of occurrences during T.

By (3), we get for the sequence g

$$(20) \qquad \varphi_g(t) = \begin{cases} \dfrac{a}{au+1}; & t \leq u \\[3mm] \dfrac{a}{au+1} e^{au} e^{-at}; & t \geq u. \end{cases}$$

Inserting (20), (15) and (14) in (7) and evaluating the integrals, we finally get

$$(21) \quad P_g(n, T) = \begin{cases} a_{n-1} - 2a_n + a_{n+1}; & n \leq \dfrac{T}{u} - 1 \\[3mm] a_{n-1} - 2a_n + (n+1) - \dfrac{aT}{au+1}; & \dfrac{T}{u} - 1 \leq n \leq \dfrac{T}{u} \\[3mm] a_{n-1} - 2\left[n - \dfrac{aT}{au+1}\right] + (n+1) - \dfrac{aT}{au+1}; & \\[3mm] & \dfrac{T}{u} \leq n \leq \dfrac{T}{u} + 1. \end{cases}$$

where

(22)
$$a_n = \frac{1}{au + 1} e^{-a(T-nu)} \sum_{v=0}^{n} \frac{(T - nu)^v a^v}{v!} (n - v), \qquad (n = 0, 1, \cdots),$$

$$a_{-1} = 0.$$

When $u = 0$, we obtain

$$a_n = e^{-aT} \sum_{v=0}^{n} \frac{T^v a^v}{v!} (n - v).$$

For the sequence f we then get the Poisson distribution

(23)
$$P_f(n, T) = \frac{(aT)^n}{n!} e^{-aT}.$$

The corresponding expression for the sequence h is much more complicated·

4. A statistical experiment. The following statistical experiment will serve as an illustration of the scheme dealt with in this paper—the transformation of a sequence and the resulting formulas, especially (21).

Groups of five figures, the last rounded up if necessary, have been extracted from tables of random sampling numbers (6). Let each group denote the first five digits for a decimal x, arbitrarily chosen between 0 and 1. The variable x is supposed to have the distribution function t for $0 \leq t < 1$. We now define a new variable, y, given by

(24)
$$y = -k \log (1 - x), \text{ [or } y = -k \log x].$$

The variable y has the distribution function given by (8), viz.

$$F(t) = 1 - e^{-at}, \text{ where } \frac{1}{a} = m = k \log e. \qquad .$$

Transforming each group, or number x, according to (24), we get a sample of consecutive distances between elements in the sequence f considered in the previous sections. Choosing a constant u, we can construct the corresponding sequence g. Beginning with a point, arbitrarily chosen on the first distance, we can finally count the number of elements in successive intervals of the same length.

Take $k = 1$, $u = 0.2$ and $T = 1.5$. We then have for the sequences f and g:

$$m_f = \frac{1}{a} = \log e = 0.4343; \qquad m_g = \frac{1}{a} + u = 0.6343;$$

$$\sigma_f = \frac{1}{a} = 0.4343; \qquad \sigma_g = \frac{1}{a} = 0.4343;$$

$$M_f(T) = \frac{T}{m_f} = 3.454. \qquad M_g(T) = \frac{T}{m_g} = 2.365.$$

The experiment yielded the following results:

For the sequence f: *For the sequence g:*

Number of elements 801.   Number of elements 555.

$\bar{m}_f = 0.450$. $\bar{m}_g = 0.648$.

In neither case is the deviation between the observed and theoretical means statistically significant. In fact we have:

$$\frac{(\bar{m}_f - m_f)\sqrt{800}}{\sigma_f} \sim 1.0; \qquad \frac{(\bar{m}_g - m_g)\sqrt{554}}{\sigma_g} \sim 0.8$$

which gives $P = 0.3$ and $P = 0.4$, respectively.

TABLE I

Nos. of intervals with n elements

n	Sequence f		Sequence g		
	Observed	Expected according to (23)	Observed	Expected according to (21)	Expected according to (23)
0	6	7.6	5	8.2	23.7
1	33	26.1	53	42.5	54.8
2	48	45.1	82	81.8	63.3
3	55	51.9	69	72.2	48.8
4	36	44.8	23	29.2⎫	28.1
5	32	31.0	6	4.8⎬	13.0
6	17	17.8	1	0.2⎭	5.0⎫
7 –	12	14.7			2.4⎭
Σ	239	239	239	238.9	239
Mean	3.331	3.454	2.310	2.36	2.31
χ^2		4.825		4.524	36.7
P		0.68		0.34	<0.001

The functions a_n in (22) can be calculated by means of Pearson's tables of the incomplete γ-function (7). In the notation of these tables we obtain

$$e^{-\lambda} \sum_{v=r}^{\infty} \frac{\lambda^v}{v!} = I\left(\frac{\lambda}{\sqrt{r-1}}; r-2\right) = I(p, q).$$

Hence

$$a_n = \frac{n}{au+1} e^{-\lambda} \frac{\lambda^n}{n!} + \frac{n-\lambda}{au+1} [1 - I(p, q)],$$

where

$$\lambda = a(T - nu); \qquad p = \frac{\lambda}{\sqrt{n-1}}; \qquad q = n - 2.$$

In the present case, however, we only need the numbers up to a_7. Accordingly, the a_n have been calculated directly.

The resulting theoretical and observed distributions for the number of elements during T for the sequences f and g will be found in Table I. For comparison, a Poisson distribution, with the same mean as observed for the sequence g_1 is given. The result of a χ^2 test is also shown in Table I. Judged by the χ^2 test the distributions (23) and (21) agree fairly well with the observed distributions. As was to be expected, the Poisson distribution cannot be used for the sequence g.

REFERENCES

[1] L. ALAOGLU AND N. M. SMITH, "Statistical theory of a scaling circuit," *Phys. Rev.*, Vol. 53 (1938), pp. 832 836.

[2] L. v. BORTKIEWICZ, *Die radioaktive Strahlung als Gegenstand wahrscheinlichkeitstheoretischer Untersuchungen*, Berlin, 1913.

[3] B. V. GNEDENKO, "On the theory of Geiger-Müller counters," (in Russian), *Jour. for Exp. and Theor. Phys.*, Vol. 11 (1941), pp. 101 106.

[4] C. LEVERT AND W. L. SCHEEN, "Probability fluctuations of discharges in a Geiger-Müller counter produced by cosmic radiation," *Physica*, Vol. 10 (1943), pp. 225-238.

[5] A. E. RUARK AND F. E. BRAMMER, "The efficiency of counters and counter circuits," *Phys. Rev.*, Vol. 52 (1937), pp. 322-324.

[6] M. G. KENDALL AND B. BABINGTON SMITH, *Tables of Random Sampling Numbers*, Tracts for Computers XXIV, Cambridge, 1939.

[7] KARL PEARSON, *Tables of the Incomplete Γ-function*, Cambridge, 1922.

THE PROBABILITY FUNCTION OF THE PRODUCT OF TWO NORMALLY DISTRIBUTED VARIABLES[1]

By Leo A. Aroian

Hunter College

1. Introduction and summary. Let x and y follow a normal bivariate probability function with means $\bar{X}$, $\bar{Y}$, standard deviations σ_1, σ_2, respectively, r the coefficient of correlation, and $\rho_1 = \bar{X}/\sigma_1$, $\rho_2 = \bar{Y}/\sigma_2$. Professor C. C. Craig [1] has found the probability function of $z = xy/\sigma_1\sigma_2$ in closed form as the difference of two integrals. For purposes of numerical computation he has expanded this result in an infinite series involving powers of z, ρ_1, ρ_2, and Bessel functions of a certain type; in addition, he has determined the moments, seminvariants, and the moment generating function of z. However for ρ_1 and ρ_2 large, as Craig points out, the series expansion converges very slowly. Even for ρ_1 and ρ_2 as small as 2, the expansion is unwieldy. We shall show that as ρ_1 and $\rho_2 \to \infty$, the probability function of z approaches a normal curve and in case $r = 0$ the Type III function and the Gram-Charlier Type A series are excellent approximations to the z distribution in the proper region. Numerical integration provides a substitute for the infinite series wherever the exact values of the probability function of z are needed. Some extensions of the main theorem are given in section 5 and a practical problem involving the probability function of z is solved.

2. Theorems on approach to normality. The moment generating function of z, $M_z(\theta)$, is [1]

$$(2.1) \qquad M_z(\theta) = \frac{\exp \dfrac{(\rho_1^2 + \rho_2^2 - 2r\rho_1\rho_2)\theta^2 + 2\rho_1\rho_2\theta}{2[1-(1+r)\theta][1 + (1 - r)\theta]}}{\sqrt{[1 - (1+r)\theta][1 + (1 - r)\theta]}}.$$

Let $\bar{z}$, and σ_z be the mean and the standard deviation of z, and $t_z = (z - \bar{z})/\sigma_z$. Now

$$(2.2) \qquad \bar{z} = \rho_1\rho_2 + r, \qquad \sigma_z = \sqrt{\rho_1^2 + \rho_2^2 + 2r\rho_1\rho_2 + 1 + r^2}.$$

Using (2.2) we find in the usual way the moment generating function of t_z

$$(2.3) \quad M_{t_z} = \frac{\exp \dfrac{-2rw + (\rho_1^2 + \rho_2^2 + 2r\rho_1\rho_2)w^2 + 4r^2w^2 - 2w^3(r^2 - 1)(\rho_1\rho_2 + r)}{2[1 - (1 + r)w][1 + (1 - r)w]}}{\sqrt{[1 - (1 + r)w][1 + (1 - r)w]}},$$

where $w = \theta/\sigma_z$.

Consider $r \geq 0$. Then in the limit as ρ_1 and $\rho_2 \to \infty$ in any manner whatever,

$$(2.4) \qquad \lim_{\rho_1, \rho_2 \to \infty} M_{t_s}(\theta) = e^{\theta^2/2},$$

and by the theorem of Curtiss [2] on moment generating functions we see in the limit as ρ_1, $\rho_2 \to \infty$ the probability function of z approaches a normal curve with mean, $\bar{z}$, and variance σ_z^2, $r \geq 0$.

In case $-1 + \epsilon < r < 0, \epsilon > 0$, some care is required wherever

$$\sqrt{\rho_1^2 + \rho_2^2 + 2\rho_1\rho_2 r}$$

occurs. If one uses $\rho_1^2 + \rho_2^2 \geq 2\rho_1\rho_2$, the proof goes forward quite readily. Hence we have proved the theorem:

THEOREM (2.5). *The distribution of z approaches normality with mean $\bar{z}$, and variance σ_z^2 as ρ_1 and $\rho_2 \to \infty$ in any manner whatever, $-1 + \epsilon < r \leq 1$, $\epsilon > 0$.*

It is evident in Theorem (2.5) we may allow ρ_1, $\rho_2 \to -\infty$ without any other changes. Theorems (2.6) and (2.7) are proved in essentially the same way as (2.5).

THEOREM (2.6). *The distribution of z approaches normality with mean $\bar{z}$, and variance σ_z^2, if $\rho_1 \to \infty, \rho_2 \to -\infty, -1 \leq r < 1 - \epsilon, \epsilon > 0$.*

THEOREM (2.7). *The distribution of z approaches normality if ρ_1 remains constant $\rho_2 \to \infty, -1 + \epsilon < r \leq 1, \epsilon > 0$; or if ρ_1 remains constant $\rho_2 \to -\infty, -1 \leq r < 1 - \epsilon, \epsilon > 0$.*

Naturally in any of the theorems ρ_1 and ρ_2 may be interchanged. In practice ρ_1 and ρ_2 are usually positive. The approach to normality is more rapid if both ρ_1 and ρ_2 have the same sign as r.

3. Numerical values. In order to show how closely the Type III and the Gram-Charlier Type A series approximate the probability function of z, $f(z)$, or more precisely $f(z, \rho_1, \rho_2, r)$, we use numerical integration where

$$f(z, \rho_1, \rho_2, r) = I_1(z) - I_2(z),$$

$$I_1(z) = \frac{1}{2\pi\sqrt{1 - r^2}} \int_0^\infty \exp - \frac{1}{2(1 - r^2)} \left\{ (x - \rho_1)^2 - 2r(x - \rho_1)\left(\frac{z}{x} - \rho_2\right) \right.$$
$$(3.1)$$
$$\left. + \left(\frac{z}{x} - \rho_2\right)^2 \right\} \frac{dx}{x},$$

and $I_2(z)$ is the integral of the same function over $(-\infty, 0)$, [1]. Now $I_1(z)$ may be written as

$$(3.2) \qquad I_1(z) = \frac{1}{\sqrt{1 - r^2}} \int_0^\infty \varphi(t_1)\varphi(t_2)\beta(t_3) \frac{dx}{x},$$

where

$$\varphi(t) = \frac{e^{-(t^2/2)}}{\sqrt{2\pi}}, \qquad t_1 = \frac{x - \rho_1}{\sqrt{1 - r^2}}, \qquad t_2 = \left(\rho_2 - \frac{z}{x}\right) \Big/ \sqrt{1 - r^2},$$

$$\beta(t_3) = e^{t_3}, \quad t_3 = r t_1 t_2.$$

We readily obtain $I_1(z)$ $\sqrt{1 - r^2}$ by forming the product of $\varphi(t_1)$, $\varphi(t_2)$, $\beta(t_3)$, and $1/x$ using numerical integration applying Weddle's formula, the Gregory-Newton formula, or the simple rectangular formula depending on circumstances. The rectangular formula [3] is remarkably accurate when the function $T = \varphi(t_1)\varphi(t_2)\beta(t_3)/x$ in the interval 0 to ∞ or 0 to $-\infty$ is somewhat symmetrical. Appropriate tables for $\varphi(t_1)$, $\varphi(t_2)$ (see [4]), $\beta(t_3)$ (see [5]) and $1/x$ (see [6]) are readily available. In the important case of the independence of x and y, $r = 0$ and (3.2) becomes

$$(3.3) \qquad I_1(z) = \int_0^\infty \varphi(t_1)\varphi(t_2) \frac{dx}{x}, \qquad t_1 = x - \rho_1, \qquad t_2 = \rho_2 - \frac{z}{x}.$$

4. Approximations to $f(z)$. When $r = 0$, the standard seminvariants ξ_3, and ξ_4 of z are

$$(4.1) \qquad \xi_3 = \frac{6\rho_1\rho_2}{(\rho_1^2 + \rho_2^2 + 1)^{3/2}}, \qquad \xi_4 = \frac{6\{2(\rho_1^2 + \rho_2^2) + 1\}}{(\rho_1^2 + \rho_2^2 + 1)^2}$$

remembering

$$\bar{z} = \rho_1\rho_2, \qquad \sigma_z = \sqrt{\rho_1^2 + \rho_2^2 + 1}.$$

In the Pearson system (see [7]) δ, the criterion, is

$$(4.2) \qquad \delta = \frac{2\xi_4 - 3\xi_3^2}{6 + \xi_4}$$

and for the probability function of z

$$(4.3) \qquad \delta = \frac{2(\rho_1^2 + \rho_2^2 + 1)\{2(\rho_1^2 + \rho_2^2) + 1\} - 18\rho_1^2\rho_2^2}{(\rho_1^2 + \rho_2^2 + 1)[(\rho_1^2 + \rho_2^2 + 1)^2 + 2(\rho_1^2 + \rho_2^2) + 1]}$$

and if $\rho_1 = \rho_2 = \rho$

$$(4.4) \qquad \delta = \frac{2(4\rho^2 + 1)(2\rho^2 + 1) - 18\rho^4}{(2\rho^2 + 1)[(2\rho^2 + 1)^2 + (4\rho^2 + 1)]}.$$

Now $\delta = 0$, $\xi_3 \neq 0$, for the Type III function, and clearly $\lim\limits_{\rho_1, \rho_2 \to \infty} \delta = 0$. By use of (3.3) the accurate values of $f(z)$ have been calculated for various combinations of ρ_1 and ρ_2 and compared with the Type III approximation using $\bar{z}$, σ_z, ξ_3.

(4.5) Investigations so far completed show that for $\rho_1 \geqq 4$ and $\rho_2 \geqq 4$ simultaneously, and $|\delta| \leqq .008$, the Type III approximation will provide values of t_s correct to three significant figures at least where

$$(4.6) \qquad \int_{-\infty}^{t_s^{(1)}} f(t_s) = \alpha, \qquad \int_{t_s^{(2)}}^\infty f(t_s) = \alpha, \quad \text{and} \quad .05 \leqq \alpha \leqq .005.$$

These are the values of t_s which would be needed in testing hypotheses. The exact values of $t_s^{(1)}$ and for $t_s^{(2)}$ for various values of ρ_1 and ρ_2 less than 4 will be

determined it is hoped in the future and will be published along with the comparisons of the Type III values of t_z with the accurate values of t_z in the important borderline cases of $\rho_1 = \rho_2 = 2$, and $\rho_1 = \rho_2 = 3$. The values of $f(z)$ for $\rho_1 = \rho_2 = 2$ and $\rho_1 = \rho_2 = 4$ have been calculated but these are being withheld for a more complete table. The table of values of $\bar{z}$, σ_z, ξ_3, ξ_4, and δ (Table II) shows then that the Type III function is excellent along a band about $\rho_1 = \rho_2$, since $\xi_3 \neq 0$, and δ is very small.

We use the Gram-Charlier Type A series of three terms to approximate the probability function of z in t_z units.

$$(4.7) \qquad f(t_z) \sim \varphi(t) - \frac{\xi_3 z}{3!} \varphi^{(3)}(t) + \frac{\xi_4 z}{4!} \varphi^{(4)}(t),$$

in the usual notation.

TABLE I

t_z	$f(t_z)$ Correct value	Normal Curve	Gram-Charlier Type A
.9950372	.2406367	.2431716	.2408235
1.4925558	.1275209	.130970	.127484
1.9900744	.0538243	.0550708	.053704
2.4875930	.0184606	.0180791	.0184500
2.9851116	.0052477	.0046338	.0052944
3.4826302	.0012609	.0009272	.0012804
3.9801488	.0002611	.0001449	.000260
4.4776674	.0000467	.0000177	.0000425
4.9751860	.00000745	.00000168	.00000555

(4.8) For $|\xi_3| < .5$ and $\xi_4 < .4$ simultaneously the Gram-Charlier Type A series is quite adequate for finding probability levels such as those of (4.6). These will in general give 3 significant figures for $t_z^{(1)}$ or $t_z^{(2)}$. In the special case $\rho_1 = 0$, $\rho_2 = 10$, the Gram-Charlier Type A series differs from $f(t_z)$ very slightly in the range $1 \leqq |t_z| < \infty$ (see Table 1). Naturally the Gram-Charlier will be used wherever Type III is not indicated, although there exist some overlapping regions where either one may be used. It should be noticed that the approach of $f(z)$ to normality is more rapid along a row than down a diagonal. In case either ρ_1 or ρ_2 is negative, we may make use of the equation

$$(4.9) \qquad f(z, -\rho_1, \rho_2, r) = f(-z, \rho_1, \rho_2, -r).$$

We note that when $r = 0$, $f(z, \rho_1, \rho_2)$ always possesses a discontinuity at $z = 0$, (see [1]). A table of $\bar{z}$, σ_z, ξ_3, ξ_4, and δ is provided for values of ρ_1 and ρ_2 from 0 to 10 inclusive.

TABLE II*

ρ₂ \ ρ₁	2	4	6	8	10
0	0 2.236068 0 2.160 .529	0 4.123106 0 .685121 .205	0 6.082762 0 .319942 .101	0 8.062258 0 .183195 .059	0 10.049876 0 .118224 .039
2	4 3 .8 1.259259 .020	8 4.582576 .498784 .557823 .056	12. 6.403124 .274256 .289114 .056	16. 8.306624 .167493 .172653 .042	20. 10.246951 .111531 .113742 .031
4		16. 5.744563 .506408 .358127 −.0084	24. 7.280110 .373206 .224279 .0049	32. 9. .263374 .147234 .014	40 10.816654 .189641 .102126 .016
6			36. 8.544004 .346314 .163258 −.0054	48. 10.049876 .28373 .118224 −.00083	60 11.704700 .224503 .087272 .0038
8				64. 11.357817 .262088 .092663 −.0034	80 12.845233 .226472 .072507 −.0015
10					100. 14.177447 .210551 .059553 −.0023

* The first value in a cell is $\bar{z}$, the second σ_3, the third ξ_3, the fourth ξ_4, the fifth δ.

5. Some extensions. We may generalize our results to any case where x and y are distributed approximately in a normal distribution such as the distribution of the product of two means, when the sizes of the samples N_1 and N_2 are large and consequently ρ_1 and ρ_2 will be large. Another example occurs if x and y each follows a Bernoulloi probability function with parameters p_1 and p_2 respectively where the number of trials in each case is large. We must warn the reader that the condition $\rho_1 \to \infty$, $\rho_2 \to \infty$ alone does not mean that the distribution of z approaches normality. Both x and y must be distributed normally.

The actual problem which gave rise to this investigation was the question of determining the sum of a great many variates [8]. Let T variates v_1, v_2, $\cdots$, v_T be given whose sum $A = \sum_{i=1}^{T} v_i$ is desired. Clearly

$$A = T\bar{V}_p , \quad \bar{V}_p = \sum_{i=1} v_i/T.$$

Now let us estimate A by $\tilde{A} = \tilde{T}_s \bar{V}_s$ where $\tilde{T}_s$ is an estimate of T and $\bar{V}_s$ is an estimate of $\bar{V}_p$. If $\sigma_{\tilde{T}_s}$ is very small, $\rho_1 = T/\sigma_{\tilde{T}_s}$ will be large and $\rho_2 = \bar{V}_p/\sigma_{\bar{V}_s}$ $= \sqrt{N}\bar{V}_p/\sigma_p$ will be very large. Assuming $\tilde{T}_s$ is distributed normally and obviously $\bar{V}_s$ is distributed normally for N large, we see by the theorems of this paper that $\tilde{A}$ will be distributed normally. Confidence limits for A may be calculated in the usual fashion as $\tilde{A} \pm \gamma\sigma_{\tilde{A}}$, where γ is determined by

$$\int_{t-\gamma}^{\infty} \varphi(t)dt = \alpha,$$

with α generally chosen as .025 or less and

$$\sigma_{\tilde{A}} = \sqrt{\tilde{T}_s^2 \sigma_{\bar{V}_s}^2 + \bar{V}_s^2 \sigma_{\tilde{T}_s}^2 + \sigma_{\bar{V}_s}^2 \sigma_{\tilde{T}_s}^2}.$$

Stratification is also possible. It is interesting to note that many functions which occur in life insurance are products. Such applications will be treated fully elsewhere. Naturally the critical region whether both tails or one tail of the distribution should be used depends on the alternatives to the hypothesis being tested.

Generalizations of the main theorem are possible for the probability function of $z = \prod_{i=1}^{r} x_i$ where x_1, x_2, $\cdots$, x_r follow a multivariate normal probability function. These will be investigated in a later paper. It may be noted that J. B. S. Haldane has investigated the distribution of a product along different lines [9].

REFERENCES

[1] Cecil C. Craig, "On the frequency function of xy," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 1–15.

[2] J. H. Curtiss, "A note on the theory of moment generating functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 430–434.

[3] A. L. O'Toole, "On the degree of approximation of certain quadrature formulas," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 143-153.

[4] ARNOLD N. LOWAN, Technical Director, *Tables of Probability Functions, Vol. II*. National Bureau of Standards, Washington, D. C.

[5] ARNOLD N. LOWAN, Technical Director, *Tables of the Exponential Function e^x*. National Bureau of Standards, Washington, D. C.

[6] ARNOLD N. LOWAN, Technical Director, *Tables of Reciprocals of Integers from 100,000 through 200,009*. Columbia Univ. Press, New York.

[7] CECIL C. CRAIG, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16–28.

[8] LEO A. AROIAN, "Some methods for the evaluation of a sum," *Amer. Stat. Ass. Jour.*, Vol. 39 (1944), pp. 511–515.

[9] J. B. S. HALDANE, "Moments of the distribution of powers and products of normal variates," *Biometrika*, Vol. 32 (1942), pp. 226–242.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

A REMARK ON CHARACTERISTIC FUNCTIONS

BY A. ZYGMUND

University of Pennsylvania

1. Let $F(x)$, $-\infty < x < +\infty$, be a distribution function, and

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x)$$

its characteristic function. It is well known that the existence of $\varphi'(0)$ does not imply the existence of the absolute moment

$$(1) \qquad \int_{-\infty}^{+\infty} |x|\, dF(x).$$

A simple example is provided by the function

$$\varphi(t) = C \sum_{n=2}^{\infty} \frac{\cos nt}{n^2 \log n},$$

where C is a positive constant. Since the series on the right differentiated term by term converges uniformly (see [1]), $\varphi'(t)$ exists (and is continuous) for all values of t, and in particular at the point $t = 0$. Obviously $\varphi(t)$ is the characteristic function of the masses $C/2n^2 \log n$ concentrated at the points $\pm n$ for $n = 2, 3, \cdots$. The constant C is such that the sum of all the masses is 1. The divergence of the series $\Sigma 1/n \log n$ implies that in this particular case the moment (1) is infinite.

In a recent paper (see [2], esp. p. 120, footnote), Fortet raises the problem of whether the existence of $\varphi'(0)$ implies the existence of the first algebraic moment

$$(2) \qquad \int_{-\infty}^{+\infty} x\, dF(x) = \lim_{X \to +\infty} \int_{-X}^{X} x\, dF(x).$$

The main purpose of this note is to show that this is so. We shall even prove a slightly more general result.

A function $\psi(t)$ defined in the neighborhood of a point t_0 is said to be *smooth* at this point if

$$\lim_{h \to +0} \frac{\psi(t_0 + h) + \psi(t_0 - h) - 2\psi(t_0)}{h} = 0.$$

Clearly, if ψ has a one-sided derivative at the point t_0, the derivative on the other side also exists and has the same value. Thus the graph of $\psi(t)$ has no angular point for $t = t_0$, and this explains the terminology. If $\psi'(t_0)$ exists and is finite, $\psi(t)$ is smooth for $t = t_0$. The converse is obviously false, since any

function whose graph is symmetric with respect to $t = t_0$ is smooth at that point.

THEOREM 1. *If the characteristic function $\varphi(t)$ is smooth at the point 0, then a necessary and sufficient condition for the existence of $\varphi'(0)$ is the existence of the moment (2). The value of (2) is $-i\varphi'(0)$.*

In particular, the existence and finiteness of $\varphi'(0)$ implies the existence of (2). That the converse is false, is obvious. For if a_0, a_1, a_2, $\cdots$ are positive numbers and $a_0 + 2a_1 + 2a_2 + \cdots = 1$, then $\psi(t) = a_0 + 2\Sigma_1^\infty a_n \cos nt$ is the characteristic function of the distribution function $F(x)$ corresponding to masses concentrated at the integer points $\pm n$ and having the values a_n there. Owing to the symmetry of the masses, the number (2) exists, and is zero even if $\varphi(t)$ is non-differentiable for $t = 0$ (we may e.g. take for $\varphi(t)$ the Weierstrass non-differentiable function $C \Sigma_1^\infty a^n \cos b^n t$, where C is a suitable constant).

PROOF. We may write

$$\varphi(t) = \int_0^\infty \cos xt \, dG(x) + i \int_0^\infty \sin xt \, dG(x) = \psi_1(t) + i\psi_2(t)$$

where

$$G(x) = F(x) - F(-x), \; H(x) = F(x) + F(-x).$$

Thus

(3) $$0 \leq |\Delta H| \leq \Delta G.$$

Since $\varphi(t)$ is smooth at the point 0, and since $\psi_1(t)$ is even, $\psi_2(t)$ odd,

$$0 = \lim_{h \to +0} \frac{\varphi(h) + \varphi(-h) - 2\varphi(0)}{h} = 2 \lim \frac{\psi_1(h) - \psi_1(0)}{h}$$

$$= -2 \lim_{h \to +0} \int_0^\infty \frac{1 - \cos hx}{h} \, dG(x)$$

so that, replacing h by $2h$,

$$\int_0^\infty \frac{\sin^2 hx}{h} \, dG(x) \to 0 \qquad\qquad \text{as } h \to 0.$$

Since the integrand is positive we obtain successively

$$\int_0^{1/h} \frac{\sin^2 hx}{h} \, dG(x) = o(1),$$

$$\int_0^{1/h} \frac{\left(\dfrac{2}{\pi} hx\right)^2}{h} \, dG(x) = o(1),$$

(4) $$\int_0^{1/h} x^2 \, dG(x) = o(h^{-1}),$$

$$\int_{1/2h}^{1/h} x^2 \, dG(x) = o(h^{-1}),$$

(5) $$\int_{1/2h}^{1/h} dG(x) = o(h).$$

Since $\psi_1(t)$ is even, the smoothness of $\varphi(t)$, and so also of $\psi_1(t)$, at the point $t = 0$ implies that $\psi_1'(0)$ exists and is zero. If $h \to +0$,

$$\frac{\psi_2(h) - \psi_2(0)}{h} = \int_0^\infty \frac{\sin xh}{h}\, dH(x) = \int_0^{1/h} + \int_{1/h}^\infty = A_h + B_h,$$

$$|B_h| \leq h^{-1} \int_{1/h}^\infty |dH| \leq h^{-1}\left(\int_{1/h}^{2/h} dG + \int_{2/h}^{4/h} dG + \int_{4/h}^{8/h} dG + \cdots\right)$$

$$= h^{-1}o(h + h/2 + h/4 + \cdots) = o(1),$$

by (3) and (5). Also

$$A_h - \int_0^{1/h} x\, dH = \int_0^{1/h}\left(\frac{\sin hx}{hx} - 1\right)x\, dH = \int_0^{1/h} O(x^2 h^2)x\, dG$$

$$= \int_0^{1/h} O(x^2 h)\, dG = o(1),$$

by (3) and (4). Thus

$$\frac{\psi_2(h) - \psi_2(0)}{h} = o(1) + \int_0^{1/h} x\, dH = o(1) + \int_{-1/h}^{1/h} x\, dF,$$

and so

$$\frac{\varphi(h) - \varphi(0)}{h} = o(1) + i\int_{-1/h}^{1/h} x\, dF.$$

It follows that the existence of (2) is equivalent to the existence of the right-hand side derivative of $\varphi(t)$ at the point $t = 0$, or, on account of smoothness, to the existence of $\varphi'(0)$. Moreover, the value of (2) is $-i\varphi'(0)$. This completes the proof of Theorem 1.

2. Suppose that a function $\psi(t)$ defined near the point t_0 satisfies for $h \to 0$ a relation

$$\psi(t_0 + h) = \alpha_0 + \alpha_1 h/1! + \cdots + \alpha_{k-1}h^{k-1}/(k-1)! + [\alpha_k + \sigma(1)]h^k/k!,$$

where $\alpha_0, \alpha_1, \cdots, \alpha_k$ are constants. Then α_k is called the *kth generalized derivative* of ψ at the point t_0. It will be denoted by $\psi_{(k)}(t_0)$. The existence and finiteness of $\psi^{(k)}(t_0)$ implies the existence of $\psi_{(k)}(t_0)$ and both numbers are equal.

Another generalization of higher derivatives is based on the consideration of the symmetric differences

$$\Delta_h\psi(t_0) = \psi(t_0 + h) - \psi(t_0 - h),$$

$$\Delta_h^2\psi(t_0) = \psi(t_0 + 2h) - 2\psi(t_0) + \psi(t_0 - 2h),$$

$$\Delta_h^3\psi(t_0) = \psi(t_0 + 3h) - 3\psi(t_0 + h) + 3\psi(t_0 - h) - \psi(t_0 - 3h).$$

If $\Delta_h^k \psi(t_0)/(2h)^k$ tends to a limit as $h \to +0$, this limit is called the kth symmetric derivative of ψ at the point t_0. We shall denote it by $D_k\psi(t_0)$. Clearly, $D_k\psi(t_0)$ exists and equals $\psi_{(k)}(t_0)$, if the latter number exists.

It is a simple matter to prove (see [3]) that if k is a positive even integer, and if the characteristic function $\varphi(t)$ has at $t = 0$ a finite symmetric derivative $D_k\varphi(0)$, then the kth moment $\int_{-\infty}^{+\infty} x^k \, dF(x)$ exists, and its value is $(-1)^{k/2} D_k\varphi(0)$. Conversely, the existence of $\int_{-\infty}^{+\infty} x^k \, dF(x)$ obviously implies (for k even) the existence and continuity of $\varphi^{(k)}(t)$ for all t, and in particular at the point $t = 0$.

In order to obtain an extension of Theorem 1 to the case of derivatives of odd order, we have to generalize the notion of smoothness. We shall say that a function $\psi(t)$ satisfies for $t = t_0$ condition S_k, $(k = 1, 2, \cdots)$, if

$$\Delta_h^{k+1}\psi(t_0) = o(h^k) \qquad \text{as } h \to +0.$$

For $k = 1$, condition S_k is identical with smoothness at t_0. Clearly, if $\psi_{(k)}(t_0)$ exists, ψ satisfies condition S_k at t_0.

THEOREM 2. *Suppose that k is a positive odd integer, and let $\varphi(t)$ be the characteristic function of a distribution function $F(x)$. If φ satisfies condition S_k at the point 0, a necessary and sufficient condition for the existence of $D_k\varphi(0)$ is the existence of the symmetric moment*

$$(6) \qquad \int_{-\infty}^{\infty} x^k \, dF(x) = \lim_{x \to +\infty} \int_{-x}^{x} x^k \, dF(x)$$

whose value is then equal to $i^{-k} D_k\varphi(0)$. In particular, the existence of $\varphi_{(k)}(0)$ implies that of (6).

The proof of Theorem 2 is analogous to that of Theorem 1. Let $G(x)$ and $H(x)$ have the same meaning as before. Since $k + 1$ is even, condition S_k at the point $t = 0$ gives

$$\Delta_h^{k+1}\varphi(0) = \int_{-\infty}^{+\infty} (e^{ixh} - e^{-ixh})^{k+1} \, dF(x) = 2^{k+1}(-1)^{(k+1)/2} \int_{-\infty}^{+\infty} (\sin xh)^{k+1} \, dF(x)$$

$$= 2^{k+1}(-1)^{(k+1)/2} \int_{0}^{\infty} (\sin xh)^{k+1} \, dG(x) = o(h^k),$$

so that

$$\int_{0}^{1/h} (\sin xh)^{k+1} \, dG(x) = o(h^k)$$

$$(7) \qquad \int_{0}^{1/h} x^{k+1} \, dG(x) = o(h^{-1})$$

$$(8) \qquad \int_{1/2h}^{1/h} dG(x) = o(h^k).$$

On the other hand,

$$i^{-k} \frac{\Delta_h^k \varphi(0)}{(2h)^k} = \int_{-\infty}^{+\infty} \left(\frac{\sin xh}{xh}\right)^k x^k \, dF(x) = \int_0^\infty \left(\frac{\sin xh}{xh}\right)^k x^k \, dH(x)$$

$$= \int_0^{1/h} + \int_{1/h}^\infty = A_h + B_h,$$

say. Here

$$|B_h| \le h^{-k} \int_{1/h}^\infty dG(x) = h^{-k} \left[\int_{1/h}^{2/h} + \int_{2/h}^{4/h} + \cdots \right]$$

$$= h^{-k} \left[o(h^k) + o\left(\frac{h}{2}\right)^k + \cdots \right] = o(1),$$

by (8). Since

$$\left(\frac{\sin u}{u}\right)^k = \{1 + O(u^2)\}^k = \{1 + O(u)\}^k = 1 + O(u)$$

for small u, we immediately obtain

$$A_h - \int_0^{1/h} x^k \, dH(x) = \int_0^{1/h} O(hx^{k+1}) \, dG(x) = o(1),$$

by (7). Collecting the results, we see that

$$i^{-k} \frac{\Delta_h^k \varphi(0)}{(2h)^k} - \int_0^{1/h} x^k \, dH(x) = i^{-k} \frac{\Delta_h^k \varphi(0)}{(2h)^k} - \int_{-1/h}^{1/h} x^k \, dF(x) = o(1),$$

which completes the proof of Theorem 2.

One more remark. By Theorem 2, the existence of the first moment is equivalent to the existence of the first symmetric derivative

$$D_{(1)}\varphi(0) = \lim_{h\to 0} [\varphi(h) - \varphi(-h)]/2h.$$

In Theorem 1 we have a corresponding result for ordinary first derivative

$$\varphi'(0) = \lim_{h\to 0} [\varphi(h) - \varphi(0)]/h.$$

There is no discrepancy here since at every point where φ is smooth the two notions of derivative are equivalent.

REFERENCES

[1] A. ZYGMUND, *Trigonometrical series*, Warszawa-Lwów, 1935, p. 108.
[2] R. FORTET, "Calcul des moments d'une fonction de répartition à partir de sa caractéristique," *Bull. des Sci. Math.*, Vol. 68 (1944), pp. 117–131.
[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946, p. 90.

A LOWER BOUND FOR THE VARIANCE OF SOME UNBIASED SEQUENTIAL ESTIMATES

BY D. BLACKWELL AND M. A. GIRSHICK

Howard University and *Bureau of the Census*

Consider a sequence of independent chance variables x_1, x_2, $\cdots$ with identical distributions determined by an unknown parameter θ. We assume that $E\, x_i = \theta$ and that $W_k = x_1 + \cdots + x_k$ is a sufficient statistic for estimating θ from $x_1, \cdots, x_k$. A sequential sampling procedure is defined by a sequence of mutually exclusive events S_k such that S_k depends only on $x_1, \cdots, x_k$ and $\Sigma\, P(S_k) = 1$. Define $W = W_k$ and $n = k$ when S_k occurs. In a previous paper by one of the authors [1] it was shown that if $S_k = W_k C(S_1 + \cdots + S_{k-1})$, (where $C(A)$ denotes the event that A does not occur), the function $V(W, n) = E(x_1 \mid W, n)$ is an unbiased estimate of θ, and $\sigma^2(V) \leq \sigma^2(x_1)$. It is the purpose of this note to obtain a lower bound for $\sigma^2(V)$. Our result is:

THEOREM I. $\sigma^2(V) \geq \dfrac{\sigma^2(x_1)}{E(n)}$.

We remark that the lower bound is actually attained in the classical case of samples of constant size N. For in this case, (see [1]), $V = E(x_1 \mid W_N) = W_N/N$. In fact we shall show that in a sense this is the only case in which the lower bound is attained.

The proof of Theorem I depends on certain properties of sums of independent chance variables. These, formulated more generally than is required for the proof of Theorem I, are given in

THEOREM II. *Let* x_1, x_2, $\cdots$ *be independent chance variables with identical distributions, having mean* θ *and variance* $\sigma^2(x_1)$. *Let furthermore* $\{S_k\}$ *be any sequential test for which* $E(n)$ *is finite. Let* $W = x_1 + \cdots + x_k$ *when* $n = k$. *Then*

(a) $\sigma^2(W - \theta n) \leq \sigma^2(x_1)\, E(n)$.

(b) *If* $\sigma^2(n)$ *is finite, the equality sign holds in* (a).

(c) $E[x_1(W - \theta n)] = \sigma^2(x_1)$.

PROOF OF (a). Write $y_i = x_i - \theta$, and define $Y = y_1 + \cdots + y_k$ when $n = k$. By definition,

$$(1) \qquad \sigma^2(W - \theta n) = \sum_{k=1}^{\infty} \int_{S_k} (y_1 + \cdots + y_k)^2 \, dP.$$

To prove (a), we must verify that the series on the right of expression (1) converges and has sum $\leq \sigma^2(x_1)E(n)$. Now

$$
\begin{aligned}
&\sum_{k=1}^{N} \int_{S_k} (y_1 + \cdots + y_k)^2 \, dP \\
(2) \quad &\leq \sum_{k=1}^{N-1} \int_{S_k} (y_1 + \cdots + y_k)^2 \, dP + \int_{n \geq N} (y_1 + \cdots + y_N)^2 \, dP \\
&= \sum_{k=1}^{N} \int_{n \geq k} y_k^2 \, dP + 2 \sum_{k=2}^{N} \int_{n \geq k} y_k(y_1 + \cdots + y_{k-1}) \, dP.
\end{aligned}
$$

Since the event $\{n \geq k\}$ is independent of y_k, each term in the second sum vanishes and the first sum becomes

$$
\begin{aligned}
\text{(3)} \quad \sum_{k=1}^{N} \int_{\{n \geq k\}} y_k^2 \, dP &= \sigma^2(x_1) \sum_{k=1}^{N} P\{n \geq k\} \\
&= \sigma^2(x_1)[P\{n = 1\} + 2P\{n = 2\} + \cdots NP\{n = N\} \\
&\qquad + NP\{n > N\}] \leq \sigma^2(x_1)E(n).
\end{aligned}
$$

This establishes Theorem II(a).

PROOF OF THEOREM II(b).   Write $z_i = |y_i|$ and let $Z = z_1 + \cdots + z_k$ when $n = k$. From (a) it follows that $\sigma^2[(Z - nE(z_i)]$ is finite. If in addition, $\sigma^2(n) < \infty$ then $E(Z^2) < \infty$. Thus the series

$$
\text{(4)} \qquad \sum_{k=1}^{\infty} \int_{S_k} (z_1 + \cdots + z_k)^2 \, dP = \sum_{1 \leq i,j \leq k < \infty} \int_{S_k} z_i z_j \, dP
$$

converges, so that the series

$$
\text{(5)} \qquad \sum_{1 \leq i,j, \leq k < \infty} \int_{S_k} y_i y_j \, dP
$$

converges absolutely.   The terms of the latter series may be arranged to yield

$$
\text{(A):} \sum_{k=1}^{\infty} \int_{S_k} (y_1 + \cdots + y_k)^2 \, dP = \sigma^2 (W - \theta n)
$$

or to yield

$$
\text{B:} \sum_{k=1}^{\infty} \int_{\{n \geq k\}} y_k^2 \, dP + 2 \sum_{k=2}^{\infty} \int_{\{n \geq k\}} y_k(y_1 + \cdots + y_{k-1}) \, dP = \sigma^2(x_1)E(n).
$$

This proves Theorem II(b).

PROOF OF THEOREM II(c).   It follows from Theorem II(a) that $Ex_1(W - \theta n)$ is finite.   If we show that

(6) $E(W - \theta n \mid x_1) = x_1 - \theta$, i.e. $E(Y \mid y_1) = y_1$, it will follow [1] that

(7) $E[x_1(W - \theta n)] = E[x_1(x_1 - \theta)] = \sigma^2(x_1)$.

To verify (6), it is sufficient to show that if $f(x_1)$ is the characteristic function of an event depending only on x_1 (i.e. $f(x_1) = 1$ when the event occurs, $f(x_1) = 0$ otherwise)

$$
\text{(8)} \qquad E(fy_1) = E(fY).
$$

Write $\phi_1 = 0, \phi_i = f \cdot (y_2 + \cdots + y_i), i \geq 2$.
Then it easily verified that

$$
\text{(9)} \qquad E(\phi_j \mid x_1, \cdots, x_i) = \phi_i \text{ for } j \geq i
$$

$$
\text{(10)} \qquad E\phi_i \leq \sum_{k=1}^{i} |y_k|
$$

$$
\text{(11)} \qquad E(\phi_i) = 0.
$$

Hence it follows [2] that $E\phi = 0$ where $\phi = \phi_i$ when $n = i$. In our case $\phi = fY - fy_1$, and $E\phi = 0$ yields (6). This completes the proof of Theorem II.

PROOF OF THEOREM I. In [1] it is proved that $E(x_1(W - \theta n)) = E[V(W - \theta n)]$. Hence employing Theorem II we get

$$(12) \qquad \sigma^2(x_1) = E[V(W - \theta n)] = \sigma(V)\sigma(W - \theta n)\rho$$

where ρ, $(0 \leq \rho \leq 1)$, is the coefficient of correlation between V and $W - \theta n$. Substituting for $\sigma(W - \theta n)$ we get

$$(13) \qquad \begin{aligned} \sigma^2(x_1) &\leq \sigma(V)\sigma(x_1)\sqrt{E(n)}\,\rho \\ &\leq \sigma(V)\sigma(x_1)\sqrt{E(n)}. \end{aligned}$$

Solving for $\sigma(V)$ we finally obtain

$$(14) \qquad \sigma^2(V) \geq \frac{\sigma^2(x_1)}{E(n)}$$

which proves Theorem I.[1]

If $\sigma^2(n)$ is finite, the equality sign in (14) will hold if and only if $\rho = 1$. We shall now prove the following.

THEOREM III. *Let N be the minimum value of n for which $P(n = N) \neq 0$. Then, a necessary and sufficient condition that $\rho = 1$ is that $P(n = N) = 1$.*

PROOF. The sufficiency of this condition follows from the fact that if $P(n = N) = 1$, $V = W/N$. To prove the necessity of this condition, we observe that if $\rho = 1$, V is a linear function of $W - n\theta$. That is,

$$(15) \qquad V = \alpha(W - n\theta) + \beta.$$

Now, since $EV = \theta$ and $E(W - n\theta) = 0$, it follows that $\beta = \theta$. Also, since by hypothesis $\sigma^2(V) = \sigma^2(x_1)/E(n)$ and $\sigma^2(W - n\theta) = \sigma^2(x_1)E(n)$, it follows that $\alpha = 1/E(n)$. Hence the estimate V is given by

$$(16) \qquad \overset{\bullet}{V} = \frac{W - n\theta}{E(n)} + \theta.$$

---

[1] Under certain regularity conditions Cramér has obtained the inequality

$$\sigma^2(x) \geq 1/E\left(\frac{\partial \log f}{\partial \theta}\right)^2$$

where $f = f(x, \theta)$ is the density function of x ([3], p. 475). Thus with the same regularity conditions, our inequality yields

$$\sigma^2(V) \geq 1/E(n)E\left(\frac{\partial \log f}{\partial \theta}\right)^2,$$

which is a special case of the results presented by J. Wolfowitz in this issue of the *Annals*.

Let N be defined as above. We note that $N < \infty$ since by hypothesis $E(n) < \infty$. Let V_N be the estimate of θ when the sequential test terminates with $n = N$. Then $V_N = W/N$. Substituting this value in (16) we get

$$(17) \qquad \frac{W}{N} - \theta = \frac{N}{E(n)}\left[\frac{W}{N} - \theta\right].$$

We exclude the trivial case where $W \equiv N\theta$. Then (16) yields $E(n) = N$. That is $P(n = N) = 1$. This proves the theorem.

We remark that N may be a function of θ but for a fixed θ, $n = N$ is fixed when $\rho = 1$.

REFERENCES

[1] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation." Submitted to *Annals of Math. Stat.*
[2] D. BLACKWELL AND M. A. GIRSHICK, "On sums of sequences of independent chance vectors, with applications to the random walk in k dimensions," *Annals of Math. Stat.*, Vol. 17 (1946).
[3] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

AN EXTENSION TO TWO POPULATIONS OF AN ANALOGUE OF STUDENT'S *t*-TEST USING THE SAMPLE RANGE

BY JOHN E. WALSH

Princeton University

1. Summary. The modified *t*-test considered by Daly[1] (see [1]) is used to develop one-sided significance tests to decide whether the mean of a new normal population exceeds the mean of an old normal population having the same variance. Significance tests are also developed to decide whether the mean of the new population is less than the mean of the old population. These tests require very little computation for their application and are approximately as powerful as the most powerful tests of these hypotheses.

2. Introduction. Let $r_1, \cdots, r_n$, $(n \leq 10)$, be independently distributed according to a normal distribution with zero mean and unit variance. Let $r_{(u)}$ denote the uth largest of the r's. Then Daly has shown how to determine numbers g_α such that

$$(1) \qquad \begin{aligned} Pr[\bar{r}/(r_{(n)} - r_{(1)}) &> g_\alpha] = \alpha \\ \cdot Pr[\bar{r}/(r_{(n)} - r_{(1)}) &< -g_\alpha] = \alpha. \end{aligned}$$

This note will use these relations to develop easily applied significance tests to decide whether the mean ν of a new normal population exceeds the mean μ of

---

[1] This problem is also considered by Lord in [2]. This note was in proof when [2] appeared.

an old normal population with the same variance. Significance tests are also developed to test $\nu < \mu$. The simplest case considered is that of testing a new sample value x on the basis of n past sample values $y_1, \cdots, y_n$. Then the significance test at significance level α to decide whether ν exceeds μ consists in accepting $\nu > \mu$ if ·

$$x > \bar{y} + g_\alpha \sqrt{n+1}[y_{(n)} - y_{(1)}],$$

where $y_{(u)}$ is the uth largest of $y_1, \cdots, y_n$.

The significance test of $\nu < \mu$ consists in accepting $\nu < \mu$ if

$$x < \bar{y} - g_\alpha \sqrt{n+1}\,[y_{(n)} - y_{(1)}].$$

These tests are generalized to the case in which x is the mean of a sample of size r from the new population, each of $y_1, \cdots, y_n$ is the mean of a sample of size s from the old population, and z is the mean of a sample of size t from the old population. Then the tests at significance level α take the form

(2)
$$\text{Accept } \nu > \mu \text{ if } x > (1 - C_1)\bar{y} + C_1 z + g_\alpha[y_{(n)} - y_{(1)}];$$
$$\text{Accept } \nu < \mu \text{ if } x < (1 - C_1)\bar{y} + C_1 z - g_\alpha[y_{(n)} - y_{(1)}],$$

where C_1 is a given constant which is selected by the person applying the test. The introduction of the terms z and C_1 allows less reliable past information to be utilized by lumping it together in the z term and using the constant C_1 to weight this information according to its relative importance with respect to the y's.

The power of test (2) is compared with that of the corresponding Student *t*-test for the case $C_1 = 0$ and $n \leq 10$. In this comparison the quantities $x, y_1, \cdots, y_n$ are considered to be the given sample values which are used for the test, that is, the quantities from which the means $x, y_1, \cdots, y_n$ were formed are not given. It is found that the power of the Student *t*-test is only slightly greater than that of the corresponding test (2). For the cases considered, however, it is well known that the most powerful test of $\nu > \mu$ using the quantities $x, y_1, \cdots, y_n$ is the appropriate Student *t*-test. Similarly for testing $\nu < \mu$. Thus the tests (2) considered are approximately as powerful as the most powerful tests of $\nu > \mu$ and $\nu < \mu$ which use $x, y_1, \cdots, y_n$.

Examination of (2) shows that the amount of computation required for the application of one of these tests is small. Consequently the tests (2) have the desirable properties of being easily computed and nearly as powerful as any tests which could be used for the given hypotheses. This suggests their use in repetitive testing procedures which are concerned with the testing of the mean of a new sample on the basis of the means of previous samples.

3. Statement of tests. In this section three significance tests of increasing generality are stated. It is to be observed that each test is a particular example of the test following it so that tests (*A*) and (*B*) are special cases of test (*C*).

The reason for stating tests (A) and (B) is that these tests have a much simpler appearance and will cover most cases of practical application.

(A). Let each of $x, y_1, \cdots, y_n$ represent the mean of a sample of size r; let the values of the sample whose mean is x have the distribution $N(\nu, \sigma^2)$ and the values of the samples whose means are $y_1, \cdots, y_n$ have distribution $N(\mu, \sigma^2)$, where the notation $N(\xi, \sigma^2)$ denotes the normal distribution with mean ξ and variance σ^2. Then the significance test of $\nu > \mu$ at significance level α is

$$Accept \quad \nu > \mu \quad if \quad x > \bar{y} + g_\alpha \sqrt{\frac{n+1}{r}} [y_{(n)} - y_{(1)}].$$

The significance test to decide whether $\nu < \mu$ is

$$Accept \quad \nu < \mu \quad if \quad x < \bar{y} - g_{(\alpha)} \sqrt{\frac{n+1}{r}} [y_{(n)} - y_{(1)}].$$

(B). Let x equal the mean of r sample values from $N(\nu, \sigma^2)$ and each of $y_1, \cdots, y_n$ equal the mean of s sample values from $N(\mu, \sigma^2)$. The significance test for $\nu > \mu$ at significance level α is

$$Accept \quad \nu > \mu \quad if \quad x > \bar{y} + g_\alpha \sqrt{\frac{n}{r} + \frac{1}{s}} [y_{(n)} - y_{(1)}].$$

The test of $\nu < \mu$ is given by

$$Accept \quad \nu < \mu \quad if \quad x < \bar{y} - g_\alpha \sqrt{\frac{n}{r} + \frac{1}{s}} [y_{(n)} - y_{(1)}].$$

(C). Let x equal the mean of r sample values from $N(\nu, \sigma^2)$, each of $y_1, \cdots, y_n$ equal the mean of a sample of size s from $N(\mu, \sigma^2)$, z equal the mean of a sample of size t from $N(\mu, \sigma^2)$, and C_1 be a given constant value. Then the significance test of $\nu > \mu$ at significance level α is

$Accept \; \nu > \mu \; if$

$$x > (1 - C_1)\bar{y} + C_1 z + [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}.$$

The significance test to decide whether $\nu < \mu$ is

$Accept \; \nu < \mu \; if$

$$x < (1 - C_1)\bar{y} + C_1 z - [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}$$

Values of g_α for $\alpha = .05$ are given in Table I. These values were listed by Daly in [1].[2]

---

[2] Values of g_α for $\alpha = .05, .025, .01, .005, .001.$ and $.0005$ are listed in Table 9 of [2] for sample sizes from 2 to 20.

4. Derivation of tests. As tests (A) and (B) are particular cases of test (C), it is sufficient to derive test (C).

<div align="center">

TABLE I

Estimated Values of $g_{.05}$

</div>

n	$g_{.05}$
3	.882
4	.526
5	.385
6	.309
7	.260
8	.227
9	.202
10	.183

Let the quantities x', y'_1, $\cdots$, y'_n, z' be defined by

$$x' = \frac{(x - \nu) \sqrt{r}}{\sigma}, \qquad y'_i = \frac{(y_i - \mu) \sqrt{s}}{\sigma}, \qquad (i = 1, \cdots, n),$$

$$z' = \frac{(z - \mu) \sqrt{t}}{\sigma}.$$

Then x', y'_1, $\cdots$, y'_n, z' are independently distributed according to $N(0, 1)$. Define

$$r_u = \frac{1}{K_1} \left(K_1 y'_u - \sum_1^n y'_i + K_2 x' + K_2 C z' \right), \qquad (u = 1, \cdots, n).$$

It is easily verified that

$$E(r_u) = 0, \qquad E(r_u^2) = \frac{1}{K_1^2} [K_1^2 + (1 + C^2)K_2^2 - 2K_1 + n]$$

$$E(r_u r_v) = \frac{1}{K_1^2} [(1 + C^2)K_2^2 - 2K_1 + n], \qquad (u \neq v).$$

Thus, if K_1 and K_2 satisfy the equations

(3)
$$\left(\sqrt{\frac{r}{s}} + C \sqrt{\frac{t}{s}} \right) K_2 + K_1 - n = 0$$

$$(1 + C^2)K_2^2 - 2K_1 + n = 0,$$

the r_u will be independent of μ when $\mu = \nu$. Also they will be independently distributed according to $N(0, 1)$.

Rewriting the r_u in terms of $x, y_1, \cdots, y_n, z$ one obtains

$$(4) \quad r_u = \frac{\sqrt{s}}{K_1 \sigma}\left[K_1 y_u - \sum_1^n y_i + K_2 \sqrt{\frac{r}{s}}\, x + K_2 C \sqrt{\frac{t}{s}}\, z + K_2 \sqrt{\frac{r}{s}}(\mu - \nu)\right].$$

Using (3) the mean of the r_u is found to be

$$\bar{r} = \frac{K_2 \sqrt{r}}{K_1 \sigma}\left[x - \left(1 + C\sqrt{\frac{t}{r}}\right)\bar{y} + C\sqrt{\frac{t}{r}}\, z - (\nu - \mu)\right].$$

Let $r_{(u)}$ denote the uth largest of $r_1, \cdots, r_n$. Then from (1)

$$\alpha = Pr[\bar{r}/(r_{(n)} - r_{(1)}) > g_\alpha] = Pr\left[\frac{K_2\sqrt{r}}{K_1}\left(x - \left(1 + C\sqrt{\frac{t}{r}}\right)\bar{y}\right.\right.$$
$$\left.\left. + C\sqrt{\frac{t}{r}}\, z - (\nu - \mu)\right) \middle/ (y_{(n)} - y_{(1)}) > g_\alpha\right]$$

It is easily proved from (3) that

$$\frac{K_1}{K_2\sqrt{r}} = \pm\sqrt{\frac{1 + C^2}{r}\left(n + \frac{(\sqrt{r} + C\sqrt{t})^2}{s(1 + C^2)}\right)}.$$

Choosing the positive sign, putting $C = -\sqrt{\frac{r}{t}}\, C_1$, and letting $\mu = \nu$ one obtains

$$Pr\left[x > (1 - C_1)\bar{y} + C_1 z\right.$$
$$\left. + [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}\right] = \alpha,$$

verifying the first part of test (C). The second part of test (C) is verified by choosing the negative sign for $\dfrac{K_1}{K_2\sqrt{r}}$ (or by repeating the above argument using the second part of (1)).

5. Power comparison with t-test. Let $x, y_1, \cdots, y_n$ satisfy the conditions of test (B) in section 3. Then Student's t using $x, y_1, \cdots, y_n$ is given by

$$t = \frac{[x - \bar{y} - (\nu - \mu)]}{\sqrt{\sum_1^n (y_i - \bar{y})^2}} \cdot \sqrt{\frac{n - 1}{s\left(\frac{1}{r} + \frac{1}{ns}\right)}}.$$

The Student t-test based on this value of t furnishes the most powerful test of $\nu > \mu$ (and $\nu < \mu$) using $x, y_1, \cdots, y_n$. The purpose of this section is to show that test (B) has approximately the same power as this Student t-test for $n \leq 10$.

Daly has shown (see [1]) that if $r_1, \cdots, r_n$ are independently distributed according to $N(\xi, \sigma^2)$, then the test based on

$$(\bar{r} - \xi)/(r_{(n)} - r_{(1)})$$

has approximately the same power for testing $\xi > 0$ (and $\xi < 0$) as the corresponding Student t-test based on

$$(5) \qquad t = \frac{(\bar{r} - \xi)\sqrt{n(n-1)}}{\sqrt{\sum_{1}^{n}(r_i - \bar{r})^2}}$$

for $n \leq 10$.

Using the notation of section 4 let

$$r_u = \frac{\sqrt{s}}{K_1}\left[K_1 y_u - \sum_{1}^{n} y_i + K_2\sqrt{\frac{r}{s}}x\right], \qquad (u = 1, \cdots, n),$$

where $\dfrac{K_1}{K_2} > 0$. Then from consideration of (4) with $C = 0$ it is seen that the r_u are independently distributed according to $N(\xi, \sigma^2)$, where ξ equals a positive constant times $(\nu - \mu)$. Following the derivations in section 4 with $C = 0$, it is seen that the test of $\xi > 0$ with this particular choice of the r_u is identical with the test of $\nu_\bullet > \mu$ given in (B) of section 3. Similarly the test of $\xi < 0$ is identical with the test (B) of $\nu < \mu$. Thus the test (B) has approximately the same power for testing $\nu > \mu$ (and $\nu < \mu$) as the Student t-test based on the value of t given in (5) if $n \geq 10$. Replacing the r_u in (5) by their values in terms of $x, y_1, \cdots, y_n, n, r$, and s, it is found that (5) becomes

$$t = \frac{[x - \bar{y} - (\nu - \mu)]}{\sqrt{\sum_{1}^{n}(y_i - \bar{y})^2}} \cdot \sqrt{\frac{n-1}{s\left(\dfrac{1}{r} + \dfrac{1}{ns}\right)}}.$$

This proves that test (B) is approximately as powerful for testing $\nu > \mu$ and $\nu < \mu$ as the most powerful test based on the quantities $x, y_1, \cdots, y_n$ if $n \leq 10$. As test (A) is a particular case of test (B), these results also apply to test (A).

REFERENCES

[1] J. F. Daly, "On the use of the sample range in an a analogue of Student's t-test". *Annals of Math. Stat*, Vol. 17 (1946), pp. 71-74.
[2] E. Lord, "The use of range in place of standard deviation in t-test," *Biometrika*, Vol. 34 (1947), pp. 41-67.

ON THE NORM OF A MATRIX

By Albert H. Bowker
University of North Carolina

In studying the convergence of iterative procedures in matrix computation and in setting limits of error after a finite number of steps, Hotelling [1] used the square root of the sum of squares of the elements of a matrix as its norm. A wide class of functions exists which may be employed as norms in matrix calculation and substituted directly in the expressions derived by Hotelling. The

purpose of this note is to make a few general remarks about this class of functions and to propose a new norm which appears to have some value in computation.

A function $\phi(A)$ of the elements of a real matrix A may be termed a legitimate norm if it has the following four properties:

(1) $\phi(cA) = |c| \phi(A)$, c a scalar;

(2) $\phi(A + B) \leq \phi(A) + \phi(B)$, if $A + B$ is defined;

(3) $\phi(AB) \leq \phi(A)\phi(B)$, if AB is defined;

(4) $\phi(e_{ij}) = 1$, where e_{ij} is a fundamental unit matrix

whose elements are all zero except the one in the ith row and jth column, whose value is unity. These four conditions are identical with the first four axioms of Rella [2], who has shown them to be independent. Properties (1), (2), and (3) are used directly in investigations of convergence and error, but the importance of property (4) is indicated by some of its immediate consequences. Clearly $e_i' A e_j = a_{ij}$, where e_i is a fundamental unit vector. From (3) and (4) it follows that $|a_{ij}| \leq \phi(A)$ for all i and j and we have that

(5) $\max_{(i,j)} |a_{ij}| \leq \phi(A)$.

Thus $\phi(A)$ has the useful property that the norm of a matrix of errors exceeds or equals the maximum possible error. Since $\phi(A^m) \leq \phi^m(A)$, it follows from (5) that the elements of A^m will tend to zero as m increases if $\phi(A) < 1$, a result which is useful in establishing convergence. Also $\phi(A) \geq 0$.

One further consequence of (1) to (4) is of interest. Suppose A is a square matrix and let λ be any of its roots. Then there exists a non-null vector x such that $Ax = \lambda x$. Now $\phi(\lambda x) = \lambda \phi(x) \leq \phi(A)\phi(x)$ and we have

(6) $\lambda \leq \phi(A)$.

Thus, every legitimate norm is an upper bound to the characteristic roots.

Clearly many functions exist which satisfy (1) to (4). The norm used by Hotelling is $N(A) = \sqrt{\sum_{i,j} a_{ij}^2}$. A new norm which may have some value is obtained as follows:

(7) $R(A) = \max_{(i)} R_i(A)$

where

$$R_i(A) = \sum_j |a_{ij}|.$$

Clearly $R(cA) = |c| R(A)$. To show that R satisfies (2), consider

$$R_i(A + B) = \sum_j |a_{ij} + b_{ij}| \leq \sum_j |a_{ij}| + \sum_j |b_{ij}| \leq R(A) + R(B).$$

Since the above inequality holds for all i,

$$R(A + B) \leq R(A) + R(B).$$

Now $AB = || \sum_\alpha a_{i\alpha} b_{\alpha j} ||$

and

$$R_i(AB) = \sum_j | \sum_\alpha a_{i\alpha} b_{\alpha j} | \leq \sum_j \sum_\alpha | a_{i\alpha} | \cdot | b_{\alpha j} |$$

$$\leq \sum_\alpha | a_{i\alpha} | R_\alpha(B) \leq R(B)R(A).$$

Hence $R(AB) \leq R(A)R(B)$. Clearly $R(e_{ij}) = 1$. Similarly it may be shown that $C(A) = \max_{(j)} \sum_i | a_{ij} |$ also satisfies the conditions of a norm.

Since the convergence of an iterative procedure is often proved by the norm being less than one, since the norm appears in the upper bound for the error after a finite number of iterations, and since the norm of a matrix of errors is taken to indicate the magnitude of the errors, a reasonable method of choosing among several available legitimate norms is to select the smallest. It is natural to inquire whether an optimum norm in this sense exists; that is, is there a function $\phi^*(A)$ such that $\phi^*(A)$ possesses properties (1) through (4) and such that $\phi^*(A) \leq \phi(A)$ for all other $\phi(A)$ satisfying these conditions. Assume such a $\phi^*(A)$ does exist. Clearly $\phi^*(A) = \phi^*(A')$, as, if either exceeded the other, the smaller could be taken as $\phi^*(A)$. Let Λ^2 be the largest root of AA'. Then by (6)

$$\Lambda^2 \leq \phi^*(AA') \leq \phi^{*2}(A) \text{ and } \Lambda \leq \phi^*(A).$$

But Rella [2] has shown that Λ possesses (1) to (4). Thus

$$\phi^*(A) = \Lambda.$$

But, for a row vector, $C(A) \leq \Lambda$. Consequently, no minimal norm exists. It is interesting to note that a worst norm does exist, namely $P(A) = \sum_{i,j} | a_{ij} |$. Since $A = \sum_{i,j} e_{ij} a_{ij}$, $\phi(A) \leq P(A)$. Clearly $P(A)$ satisfies (1) to (4) and hence is the worst possible legitimate norm.

In practical computation, the choice so far is between $N(A)$ and $R(A)$ (or $C(A)$). No general inequalities exist and it would probably be advisable to compute both. $R(A)$ may be less than $N(A)$ and indicate convergence when $N(A)$ fails to do so. Often $R(A)$ may be computed visually and convergence proved without computing the sum of squares of the elements.

The functions $N(A)$ and $R(A)$ may also be useful in finding a simple first approximation to A^{-1}. A sufficient condition that Hotelling's iterative method for finding the inverse of a matrix A will converge is that the roots of $D = 1 - AC_0$ be less than one in absolute value where C_0 is a first approximation to A^{-1}. If the iterative procedure is to be carried out by a fully automatic computing machine such as the one described by Alt [3] it may be advisable to start with a rather poor first approximation which is easy to construct. If A has positive roots and if M is any upper bound to these roots and if C_0 is a matrix with diagonal elements equal to $1/M$ and zeros elsewhere, the iterative procedure will converge but the norm of D will not necessarily be less than one. From (6), any legitimate norm may be taken as M.

Finally, it is interesting to point out the relation of this note to some work on the problem of finding upper bounds to the roots. In fact, the inequalities $\lambda \leqq N(A)$ and $\lambda \leqq R(A)$, which are consequences of (6), are Theorem 2 of Farnell [4] and Theorem 3 of Barankin [5] respectively.

REFERENCES

[1] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 1–34.

[2] T. RELLA, "Über den absoluten Betrag von Matrizen," International Congress of Mathematicians at Oslo (1936).

[3] FRANZ L. ALT, "Multiplication of matrices," *Math. Tables and Aids to Compution*, Vol. 2 (1946), pp. 12–13.

[4] A. B. FARNELL, "Limits for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 50 (1944), pp. 789–794.

[5] EDWARD W. BARANKIN, "Bounds for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 767–770.

DEFINITION OF THE PROBABLE DEVIATION

BY M. FRÉCHET

Faculty of Science, University of Paris

The probable deviation has recently been defined by E. J. Gumbel [1], [2] as the smallest of the intervals corresponding to the probability $\frac{1}{2}$. It so happened that the author was led to an equivalent definition starting from a general idea which may be applied to absolutely general cases and which, for this reason, might be of interest.

In recent years, the author has been occupied with a study of random elements of any nature (curves, surfaces, functions, qualitative elements), a study whose future seems promising, [3]. I gave a definition of the mean of such an element expressed by an abstract integral which, however, is only defined if the random element is situated in a metric vectorial (Wiener-Banach) space.[1] But[2] a still more general definition is valid if the random element is placed in any metric space. It consists of taking, as mean position of the random element X, a fixed (non-statistical) element $b = \bar{X}$ such that the function of a which represents the mean $M(X, a)^2$ of the squared distance of X to the fixed element a, is minimum for $a = b$. (In the case where X and a are numbers, and where $M(X)^2$ is finite, we know that this minimum is reached and that there is one, and only one, determination b of a). This definition has the advantage of also defining the equiprobable position of X. This is a fixed element $c = \bar{\bar{X}}$ such that $M(X, a)$ is minimum for $c = a$. (If X and a are numbers, we know that this minimum is still reached, but may be so reached by several values of $\bar{\bar{X}}$).

Since reading Gumbel's paper, a still more general definition suggested itself.

---

[1] For the definition of metric vectorial spaces see [4].

[2] See Note 2, p. 503 of [4].

The expressions $M(X, a)$ and $\sqrt{M(X, a)^2}$ themselves may be considered as distances, but as distances of two random elements *taken together*. To each of these distances corresponds as minimum, when a varies, a different "typical" function $\hat{X}$ or $\bar{X}$ $\cdots$. Thus, without supposing anything about the space into which the different trials place X, we assume that we have defined a "deviation" of two random elements X, Y taken together. We represent this function of two random variables by $(([X], [Y]))$, a notation which differs from the representation of the distance (X, Y) of the two positions X and Y with respect to a single trial. The lower boundary of the deviation $(([X], [a]))$, a function of a, which is reached for $a = \widehat{X}$ defines a "typical" position $\widehat{X}$. Moreover, the value of this $(([X], [\widehat{X}]))$ may be considered as a measure or, at least, as a numerical ranging point of the dispersion of X.

Let us abandon these generalities. They hold especially if the element X is a real valued random variable. Among the possible and reasonable[3] expressions for the deviation $(([X], [a]))$ of the numerical variate X from a fixed number a, we may use the equiprobable value of $\mid X - a \mid$ which may be called the equiprobable deviation of X from a. Thus we have, on one side, a new "typical value" of X which will be a value of a such that the equiprobable deviation of X from a is minimum, and a new measure of dispersion which is the value of this minimum and which might be called simply the equiprobable deviation of X.

In the case where X has everywhere a continuous and finite density of probability $w(X)$ we find, as typical value, what Gumbel calls the "midvalue" and represents by ξ, and, as equiprobable deviation, what Gumbel calls the "probable deviation" and represents by ζ.

We may also consider the discontinuous case, which was given as a problem to candidates of the "Certificat d'Etudes Supérieures de Calcul des Probabilités, Option Statistique Mathématique, Session May-June, 1944." They had to solve various questions of which I cite the beginning below:

"Consider n real numbers $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ and represent, by E_a, a median value of the deviations $\mid x_k - a \mid$ of the numbers x_k and a. If a varies, E_a has a minimum E which is reached by one or several values A of a.

1) Explain, in a few words, the meaning of the values E and A.

2) For simplicity's sake, suppose that n is odd $(n = 2r + 1)$. How should E and A be calculated practically? (To find the answer, investigate first how E_a varies if a varies only slightly).

3) In the case where $n = 4s + 3$ (s is an integer equal to, or larger than, zero) show that $E \leqq \dfrac{(q_3 - q_1)}{2}$

where $\qquad\qquad q_1 = x_{s+1}, \quad q_3 = x_{n-s}$."

The study of this new typical value and of this new equiprobable deviation has the advantage that their determination is very rapid and requires hardly

---

[3] See the *Remark* at end of note.

290 M. FRÉCHET

any calculations. However, we have to note an important inferiority of the equiprobable deviation of X compared to the mean and the standard deviations of X. If one or the other of the last two deviations is zero, X is a fixed number (except for the case of the probability zero). This property seems requested by the intuitive meaning which we attribute to the dispersion, and to every measure or any mark of it. Now, the equiprobable deviation lacks this property. If, for instance, X has only three values: 0, 2, 1, the first two with the probability 0.249, and the last with the probability 0.502, the equiprobable deviation of X will be zero, whereas X will be equal to its typical value 1 only with a probability of 0.502, and not with a probability equal to unity. The same holds for any distribution for which there is a point with probability exceeding $\frac{1}{2}$.

Remark. The definitions of the mean and of the equiprobable position become meaningless in the case that $M(X, a)$, or $M(X, a)^2$, is infinite. However, we succeeded in surmounting the difficulty, and to reach definitions which are valid even in this case. If X is a number, the new definitions become equivalent to the classical definitions of the mean and equiprobable value. The proofs are given in two recent articles [5], [6].

REFERENCES

[1] E. J. GUMBEL, "Definition of the probable error," *Annals of Math. Stat*, Vol. 13 (1942)' p. 110–111.
[2] E. J. GUMBEL, "Probable deviation," *Stat. Jour*, City College, (N Y), Vol. 6 (1943), pp. 25–26.
[3] M. FRÉCHET, "L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque," *Revue Scientifique*, Vol. 82 (1944), pp. 483–512.
[4] M. FRÉCHET, *Les Espaces Abstraites*, Gauthier-Villars, Paris, 1928, pp. 125–141.
[5] M. FRÉCHET, "Les éléments aléatoires de nature quelconque," *Ann. Inst. H. Poincaré*, 1947.
[6] M. FRÉCHET, "Nouvelles définitions de la valeur moyenne et des valeurs probables d'un nombre aléatoire," *Ann. Univ. de Lyon*, 1947.

THE GENERAL RELATION BETWEEN THE MEAN AND THE MODE FOR A DISCONTINUOUS VARIATE

BY M. FRÉCHET

Faculty of Science, University of Paris

Dr. Gumbel has pointed out that one of the author's arguments employed in several particular cases (see [1]) can be employed in a general case which includes them and leads to the following result: If a statistical variate R has only positive entire values differing from zero, and if its mean value $\bar{R}$ is smaller than, or equal to, unity, the same holds for its equiprobable value $\bar{\bar{R}}$ and its mode $\bar{R}$. There are two generalizations of this result which might be of interest:

1) On the one hand, the author has shown [2] that, if a variate R can only have values (entire or not) equal to, or larger than, zero, its equiprobable value

R is, at most, equal to twice its mean value $\bar{R}$, and the inequality $\overline{R}/\bar{R} \leq 2$ cannot be improved which means that the upper boundary of the first member is exactly equal to (and not less than) two. The equality is reached when R has only two values of equal probability, one of them being zero.

2) On the other hand, if R is an integer positive variate equal to, or larger than zero, it can be proven that, if $\bar{R} \leq \alpha$, we have

(1)
$$\hat{R} \leq \frac{\alpha(\alpha + 3)}{2}.$$

Here, $\bar{R}$ and $\hat{R}$ stand for the mean and for the mode of R respectively, and α is a positive integer differing from zero. For example: if R is the number of repetitions of an event with probability p, we have, for n trials, $\bar{R} = np$, whence, if α is the first integer number equal to, or larger than, $\bar{R}$ we have the inequality (1) for the most probable number of repetitions. Naturally, this inequality only has an interest if the second member of (1) is smaller than n which means that

$$\alpha(\alpha + 3) < 2n.$$

This presupposes

$$2n > np(np + 3)$$

or

$$n < \frac{2 - 3p}{p^2}$$

and, since n must be positive,

$$p < \tfrac{2}{3}.$$

To prove the inequality (1), let us write ω_ν for the probability that $R = \nu$. We have

$$\sum_0^\infty \omega_\nu = 1; \qquad \sum_0^\infty \nu\omega_\nu = \bar{R} \leq \alpha$$

whence

(2)
$$\sum_0^{\alpha-1} (\alpha - \nu)\omega_\nu \geq \sum_{\alpha+1}^\infty (\nu - \alpha)\omega_\nu.$$

Let the mode be

$$\hat{R} = \beta$$

then

$$\omega_\beta \geq \omega_\nu \; ; \; \nu = 0, 1, 2, \cdots$$

and the first member in (2) is bounded by

(3)
$$\frac{\alpha(\alpha + 1)}{2} \omega_\beta \geq \sum_0^{\alpha-1} (\alpha - \nu)\omega_\nu.$$

Now, either $\alpha < \beta$ or $\beta \leq \alpha$. In the first case the second member in (2) leads to

$$(4) \qquad \sum_{\alpha+1}^{\infty} (\nu - \alpha)\omega_\nu \geq (\beta - \alpha)\omega_\beta$$

since the second member in (4) is one of the terms occurring in the sum. The same inequality holds in the second case, $\beta \leq \alpha$, hence it holds generally. It follows from (2), (3), and (4) that

$$\frac{\alpha(\alpha + 1)}{2} \omega_\beta \geq (\beta - \alpha)\omega_\beta .$$

The probability ω_β is certainly different from zero, since $\sum_{0}^{\infty} \omega_\nu = 1$. Consequently

$$\beta - \alpha \leq \frac{\alpha(\alpha + 1)}{2}$$

or

$$\beta \leq \frac{\alpha(\alpha + 3)}{2}$$

as stated in (1).

The equality in (1) is possible only if, from (3),

$$\alpha(\omega_\beta - \omega_0) + (\alpha - 1)(\omega_\beta - \omega_1) + \cdots + (\omega_\beta - \omega_{\alpha-1}) = 0$$

and from (4)

$$\omega_{\alpha+1} + 2\omega_{\alpha+2} + \cdots + (\beta - \alpha)\omega_\beta + \cdots = (\beta - \alpha)\omega_\beta$$

whence

$$(5) \qquad \omega_0 = \omega_1 = \cdots = \omega_\beta = \cdots = \omega_{\alpha-1}$$

and

$$(5') \qquad \omega_{\alpha+1} = \omega_{\alpha+2} = \cdots = 0.$$

The existence of the exceptional case proves that the inequality (1) cannot be improved by replacing the second member by a smaller function of α. In the exceptional case, the only possible values of R are

$$R = 0, 1, 2, \cdots, \alpha - 1, \alpha, \beta,$$

and all values, except perhaps α, are equiprobable. The probability ω_α may be, but need not be, equal to ω_β.

Moreover

$$(6) \qquad \beta = \frac{\alpha(\alpha + 3)}{2} \geq \alpha$$

and $\beta = \alpha$ is possible only if $\alpha = \beta = 0$ whence, from (5), $\omega_\nu = 0$ except for $\nu = 0$ which means that R only has one value equal to zero. Except for this trivial case, we have in the exceptional case $\beta > \alpha$, and there are $\alpha + 2$ possible values for R. Then we must have

$$\omega_\beta \geqq \omega_\alpha; \qquad \sum_0^\alpha \omega_\nu + \omega_{\beta\cdot} = 1$$

whence

$$(\alpha + 1)\omega_\beta + \omega_\alpha = 1$$

and, from (5),

$$\alpha \geqq R = \omega_\beta \sum_1^{\alpha-1} \nu + \beta\omega_\beta + \alpha\omega_\alpha = \omega_\beta \left(\frac{\alpha(\alpha-1)}{2} + \frac{\alpha(\alpha+3)}{2} \right) + \alpha\omega_\alpha$$

$$= \alpha((\alpha + 1)\omega_\beta + \omega_\alpha)$$

whence

(7) $$R = \alpha.$$

From

$$1 = (\alpha + 1)\omega_\beta + \omega_\alpha \geqq (\alpha + 2)\omega_\alpha$$

follows

(8) $$\omega_\alpha \leqq \frac{1}{\alpha + 2}; \qquad \omega_\beta = \frac{1 - \omega_\alpha}{\alpha + 1}.$$

These conditions (5), (5′), and (7) are necessary and sufficient for the existence of the exceptional case.

If the equality in (1) is excluded, the mode β and the smallest integer number α which is equal to, or larger than, the mean, are related by

(9) $$\beta \leqq \frac{\alpha(\alpha + 3)}{2} - 1 = \frac{\alpha^2 + 3\alpha + 2}{2}.$$

As shown before, this general inequality, valid for any discontinuous variate, which can assume only non-negative integer values, cannot be improved without assuming specific properties of the distribution.

REFERENCES

[1] M. Fréchet, *Les probabilités associées à un système d'evénements compatibles et dependents*, Hermann et Cie., 1943, Part II, p. 5.
[2] M. Fréchet, "Comparaison de diverses mesures de la dispersion", *Rev. de l'Inst. International de Stat.*, Vol. 8 (1940), p. 5.

NOTE ON DIFFERENTIATION UNDER THE EXPECTATION SIGN IN THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS

By T. E. HARRIS

Princeton University

Let z be any chance variable and z_1, z_2, z_3, $\cdots$ a sequence of independent chance variables, each with the same distribution as z. Let $Z_N = z_1 + z_2 + \cdots + z_N$. Let $\phi(t) = Ee^{zt}$ for all complex t for which the latter exists. Let S_1, S_2, $\cdots$ be a sequence of mutually exclusive events such that S_j depends only on z_1, z_2, $\cdots$, z_j, and $\sum_{j=1}^{\infty} P(S_j) = 1$. Let the chance variable n be defined as $n = j$ when S_j occurs. Blackwell and Girshick [1], generalizing a result of Wald [2], showed that if there is a positive constant M such that

$$(1) \qquad\qquad | Z_N | < M \text{ when } n > N$$

then the identity

$$(2) \qquad\qquad E\{e^{z_n t}(\phi(t))^{-n}\} = 1$$

holds for all complex t for which $\phi(t)$ exists and $| \phi(t) | \geq 1$. Wald [3] established conditions, including the existence of $\phi(t)$ for all real t, under which (2) may be differentiated under the expectation sign an unlimited number of times.

Without assuming the existence of $\phi(t)$ for a real t-interval the following result holds: *If (1) is true and if $E(z^k)$ and $E(n^k)$ are both finite, k a positive integer, then*

$$(3) \qquad\qquad E\left\{\frac{d^k}{ds^k}\,[e^{z_n is}(\phi(is))^{-n}]_{s=0}\right\} = 0$$

where $i = \sqrt{-1}$ and s is real. Certain identities, obtained by differentiating (2) and putting $t = 0$, can also be obtained from (3). For example, if $En = 0$, and if En^2 and Ez both exist then $EZ_n^2 = Ez^2En$.

Let $P_N = P(n \leq N)$; $p_N = P(n = N)$. Let $H(j, Z_j)$ and $F(N, Z_N)$ be the conditional cumulatives of Z_j and Z_N for $n = j$ and $n > N$ respectively. Now (2) was derived by Wald [2], p. 285, from a relation, valid whenever $\phi(t)$ exists, which in the present notation becomes

$$(4) \quad \sum_{j=1}^{N} p_j \int_{-\infty}^{\infty} (\phi(t))^{-j}\, e^{z_j t}\, dH(j, Z_j) + \frac{(1 - P_N)}{(\phi(t))^N} \int_{-\infty}^{\infty} e^{z_N t}\, dF(N, Z_N) = 1.$$

Examination of Wald's derivation of (4) shows it to be valid under the present hypotheses. Now the finiteness of $E(z^k)$ clearly implies that of $E(Z_j^k \mid n = j)$. Also, since $F(N, Z_N)$ is constant outside the interval $[-M, M]$, the integral $\int_{-\infty}^{\infty} Z_N^k\, dF(N, Z_N)$ is finite. Hence we may set $t = is$ in (4) and differentiate

k times, obtaining for all real s

(5)
$$\sum_{j=1}^{N} p_j \int_{-\infty}^{\infty} \frac{d^k}{ds^k} [(\phi(is))^{-j} e^{Z_j is}] \, dH(j, Z_j)$$
$$+ (1 - P_N) \sum_{r=0}^{k} \binom{k}{r} \frac{d^r}{ds^r} [(\phi(is))^{-N}] \cdot \int_{-\infty}^{\infty} (iZ_N)^{k-r} e^{Z_N is} \, dF(N, Z_N) = 0.$$

The derivatives of $(\phi(is))^{-N}$ are sums of terms of the form $Q(N) \cdot (\phi(is))^{-N-r}$ times terms independent of N, where $Q(N)$ is a polynomial in N of degree $\leq k$. For any $r \leq k$,

$$\lim_{N \to \infty} | (1 - P_N) N^r | = \lim_{N \to \infty} \left| N^r \sum_{j=N+1}^{\infty} p_j \right| \leq \lim_{N \to \infty} \left| \sum_{j=N+1}^{\infty} j^k p_j \right| = 0,$$

since En^k is finite. Hence $\lim (1 - P_N)Q(N) = 0$. Because of (1) the integrals in the second term of (5) are bounded as $N \to \infty$. Now set $s = 0$ in (5) and then let $N \to \infty$. Since $\phi(0) = 1$, the second term of (5) approaches 0 and the limit of the first term is just the left side of (3).

For the case of a Wald sequential process, Stein [4] has shown that all moments of n are finite. In this case (3) holds whenever Ez^k is finite.

REFERENCES

[1] DAVID BLACKWELL AND M. A. GIRSHICK, "On functions of sequences of independent chance vectors, with applications to the problem of the random walk in k dimensions," *Annals of Math. Stat.*, Vol. 17 (1946), p. 310.

[2] ABRAHAM WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), p. 283.

[3] ABRAHAM WALD, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946), p. 493.

[4] CHARLES STEIN, "A note on cumulative sums," *Annals of Math. Stat.*, Vol. 17 (1946), p. 498

A UNIQUENESS THEOREM FOR UNBIASED SEQUENTIAL BINOMIAL ESTIMATION

By L. J. SAVAGE[1]

University of Chicago

In a recent note [1], J. Wolfowitz extended some of the results of a paper by Girshick, Mosteller and Savage [2] on sequential binomial estimation. The present note carries one of Wolfowitz's ideas somewhat further. The nomenclature of [1] and [2] will be used freely. The concept of "doubly simple region" introduced in [1] and assumed there only in the hypothesis of Theorem 3, will here be shown to be unnecessarily restrictive. In so doing, we find that sim-

plicity is not only a necessary (cf. Theorem 4 of [2]) but also a sufficient condition that $\hat{p}$ be the unique unbiased estimate of p for a closed region.

LEMMA. *If R is simple there is at most one bounded unbiased estimate of any given function of p.*

PROOF. If the lemma were false, there would be a non-trivial bounded unbiased estimate of zero, i.e., $m(\alpha)$ such that $| m(\alpha) |$ is bounded by a constant m^*, $m(\alpha)$ not identically zero and $E(m(\alpha) \mid p) \equiv 0$.

$$(1) \qquad\qquad E(m(\alpha) \mid p) = \sum m(\alpha)k(\alpha)p^y q^x = 0.$$

and $m(\alpha)$ not identically zero. Since R is simple we may assume (much as in the proof of Theorem 6 of [2]) that we have a boundary point such that $m(\alpha_0) \neq 0$, α_0 is below all accessible points of its own index and also below every other α for which $m(\alpha) \neq 0$. Therefore

$$(2) \qquad | m(\alpha_0) | k(\alpha_0)p^{y_0} q^{x_0} = | \sum_{y > y_0} m(\alpha)k(\alpha)p^y q^x | \leq m^* \sum_{y > y_0} k(\alpha)p^y q^x.$$

Let M denote the set of all accessible points and boundary points at which $x < x_0$ and $y = y_0 + 1$. There are at most x_0 points in M, say $\beta_1, \cdots, \beta_n$. Considering the way in which α_0 has been chosen, every path from $(0, 0)$ to an α for which $y > y_0$ passes through or to at least one point of M. Therefore when $y > y_0$

$$P(\alpha) = k(\alpha)p^y q^x = P(\alpha \mid M)P(M)$$

$$(3) \qquad\qquad\qquad\qquad \leq P(\alpha \mid M) \sum_1^n k(\beta_j)p^{y_0+1} q^{x_j},$$

$$\leq p^{y_0+1} \sum_1^n k(\beta_j)P(\alpha \mid M).$$

From inequalities (2) and (3).

$$| m(\alpha_0) | k(\alpha_0)p^{y_0} q^{x_0} \leq m^* p^{y_0+1} \left\{ \sum_1^n k(\beta_j) \right\} \sum_{y > y_0} P(\alpha \mid M)$$

$$(4)$$

$$\leq m^* p^{y_0+1} \sum_1^n k(\beta_j).$$

But it is impossible that (4) should be satisfied for small p.

Combining the Lemma with Theorem 4 of [2] we have the

THEOREM. *A necessary and sufficient condition that $\hat{p}(\alpha)$ be the unique proper (bounded) and unbiased estimate of p for a closed region R is that R be simple.*

The sufficiency part of this Theorem extends Theorem 3 of [1] from doubly simple regions to simple regions.

The author is indebted to J. Wolfowitz for his valuable suggestions in connection with the present note.

REFERENCES

[1] J. WOLFOWITZ, "On sequential binomial estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 489-493.
[2] M. A. GIRSHICK, FREDERICK MOSTELLER, and L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems with applications." *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13-23.

ACKNOWLEDGEMENT OF PRIORITY

BY H. E. ROBBINS

University of North Carolina

At the time of publication of my papers on the measure of a random set (*Annals of Math. Stat.*, Vol. 15 (1944), pp. 70–74; Vol. 16 (1945), pp. 342–347), I was unaware that the theorem on page 72 of the first paper, which affords a means of computing the expected value of the measure, had already been found by A. Kolmogoroff. (*Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Berlin, 1933, p. 41). I wish to take this opportunity of acknowledging Kolmogoroff's priority, which was pointed out by Prof. Henry Scheffé.

ABSTRACTS OF PAPERS

1. A Test of Significance of the Coefficient of Rank Correlation for more than Thirty Ranked Items. NILAN NORRIS, Hunter College.

Hotelling and Pabst (*Annals of Math. Stat.*, Vol. 7 (1936), p. 37) have suggested the use of the Tchebycheff inequality as an approximation for testing the significance of the coefficient of rank correlation in cases where the number of ranked items is too large to enable exact probabilities to be computed directly. A table prepared in accordance with this suggestion indicates that for values of the coefficient of rank correlation larger than .50 there is a wide range of corresponding numbers of ranked items greater than thirty for which at least the five per cent level of significance is satisfied.

For certain types of applications the conservativeness of the Tchebycheff test may be a virtue rather than a limitation.

2. A Generalized T Measure of Multivariate Dispersion. HAROLD HOTELLING, University of North Carolina.

The problem of combining errors in two or more dimensions to measure the accuracy of firing and bombing is similar to problems occurring in industrial quality control where different measures of quality are applied to the same article, and to problems in mental testing and other fields. If the covariances were known a priori, the solution optimum in certain senses, for a multivariate normal distribution, would be the use of $\chi^2 = \Sigma\Sigma\lambda_{ij}x_ix_j$, where $[\lambda_{ij}]^{-1}$ is the covariance matrix and x_i is the deviation in the ith dimension. Since the covariances must in all known practical cases be estimated from a preliminary sample with (say) n degrees of freedom, χ^2 may be replaced by $T^2 = \Sigma l_{ij}x_ix_j$, where $[l_{ij}]^{-1}$ is the estimated covariance matrix. This is the same T introduced by the author in 1931 as a generalization of the Student ratio t, and has the same distribution. Upon adding together the values of T^2 for different cases (e.g. for different bombs dropped with the same bombsight), a combined measure T_0^2 of over-all excellence (e.g. of the bombsight), is obtained. T_0^2 like χ^2, can be broken down into components meaningful with respect to the causal system, specifically in relation to possible sources of excessive discrepancy. Thus, if $\bar{x}_i$ is the ith coordinate of the centroid, or mean point of impact, of m bombs, we may write $T_M^2 = \Sigma\Sigma l_{ij}\bar{x}_i\bar{x}_j$, $T_D^2 = T_0^2 - T_M^2$. Then T_D is a function only of deviations from the mean point of impact. Asymptotically (for large n), T_0, T_M and T_D have the χ distribution with m, 2 and $m - 2$ degrees of freedom respectively. But the untrustworthiness of the χ distribution as an approximation is evident even with n as large as 256, for which case calculations have been made. The exact distributions of T_0 and T_D are ascertained when the number of variates p is 2, and the probability integrals are expressed as linear functions of two incomplete beta functions. In fact, T_0^2/M equals the sum of the roots of a determinantal equation of the form $|A - \lambda B| = 0$, where A and B are sample covariance matrices with n and m degrees of freedom respectively, and a similar relation holds for T_D^2 with m replaced by $m - 2$. T_0 and T_M have the distribution published in 1931, with probability integral expressible in terms of a single incomplete beta function or the variance ratio distribution. It is shown that such parameters as the circular mean deviation are best estimated with the help of the T measures, not directly by averaging individual circular deviations.

3. Asymptotic Properties of Maximum and Quasi-Maximum Likelihood Estimates. HERMAN RUBIN, Cowles Commission for Research in Economics.

The results of J. L. Doob (*Trans. Am. Math. Soc.*, Vol. 36 (1934), pp. 759–775) on consistency of maximum likelihood estimates, are generalized and extended to arbitrary measure spaces. In some special cases, results on asymptotic normality of maximum likelihood estimates can be generalized to quasi-maximum likelihood estimates (estimates based on the assumption of a likelihood function which need not be the true function).

4. The Asymptotic Distribution of the Range. E. J. GUMBEL, Newark College of Engineering.

The asymptotic distribution of the range w for initial unlimited distributions of the exponential type is obtained by convolution of the asymptotic distributions of the two extremes. Let α and u be the parameters of the distributions of the extremes for a symmetrical variate, and let $R = \alpha(w - 2u)$ be the reduced range. Then the probability $\Psi(R)$ of the reduced range is subject to the differential equation $\Psi'' + \Psi' - \Psi \exp(-R) = 0$ which may be transformed into Bessel's equation of the first order by the substitutions $R = 2(\log 2 - \log z)$, and $\Psi = zU$ The solution is $\Psi(R) = zK_1(z)$ for the asymptotic probability, and $\psi(R) = (z^2/2)K_0(z)$ for the asymptotic distribution, $K_0(z)$ and $K_1(z)$ being the modified Bessel function of the second kind of orders zero and unity. Thus tables of $\Psi(R)$ and $\psi(R)$ may be calculated for any symmetrical distribution of the exponential type. The distribution of the range w for normal samples of size 10 is already very close to the asymptotic distribution provided that the parameters α and u are determined from the mean and the standard deviation of the range. This method permits the calculation of the distribution of the range for normal samples of any size larger than 10.

5. The Corner Test for Association. JOHN W. TUKEY, Princeton University, and PAUL S. OLMSTEAD, Bell Telephone Laboratories.

Construction. In a scatter diagram, draw the two medians, that is, the median of the x values without regard to the values of y, and the median of the y values without regard to the values of x. Think of the four quadrants thus formed as being labelled $+, -, +, -$ in order, so that the two positive quadrants lie along one diagonal and the two negative along the other. Beginning at the right-hand side of the diagram, count in along the observations until forced to cross the horizontal median. Write down the number of observations met before this crossing, attaching the sign, $+$, if they lay in the $+$ quadrant, and the sign, $-$, if they lay in the $-$ quadrant. Repeat this process, moving up from below, moving to the right from the left, and moving down from above. The quantity to be used in the test is the algebraic sum of the four numbers thus written down.

Distribution. The exact distribution of this quantity when no association is present and no two x's and no two y's are alike is almost independent of sample size over the range of values where it is apt to be used. For example, a sum of 9 or more is expected less than one time in ten for all samples of size 6 or more; a sum of 15 or more, less than one time in 100 for all samples of size 10 or more; and a sum of 21 or more, less than one time in 1000 for all samples of size 14 or more. Even for infinite sample size, the sums for these fractions become only 9, 14, and 19, respectively.

Extensions. The same ideas that underlie the outside corner test for two variables may be extended in several ways to give tests for various types of association among three or more variables.

6. Consistent Estimates Based on Partially Consistent Observations, with Particular Reference to Structural Relations. J. NEYMAN AND ELIZABETH L. SCOTT, University of California.

Let $\{X_n\}$ be a sequence of independent random variables and let F_i denote the distribution of X_i. Each distribution F_i is assumed to depend on unknown parameters. If a parameter θ appears in an infinity of distributions F_i, it is called *structural*. Otherwise, it is *incidental*. The sequence $\{X_n\}$ is called *consistent* if $\{F_n\}$ has no incidental parameters. $\{X_n\}$ is called *partially consistent* if $\{F_n\}$ has both structural and incidental parameters.—Problem of fitting a straight line when both variables are subject to errors is that of a partially consistent series of observations. Let ξ and $\eta = \alpha + \beta\xi$ be two linearly connected quantities, perhaps related to particular stars, where α and β are unknown. The values ξ_i and η_i corresponding to the ith star, $(i - 1, 2, \cdots, s)$, are unknown. The observations provide measurements x_{ij} of ξ_i, $(j - 1, 2, \cdots, m_i)$, and measurements y_{ik}, $(k = 1, 2, \cdots, n_i)$, of η_i. Both m_i and n_i are bounded and small. On the other hand, s may be considered as increasing without limit.—Assume that the x_{ij} and the y_{ik} are normally distributed with variances σ_1^2 and σ_2^2 and means ξ_i and η_i respectively. Then the totality of observations will form a partially consistent system with the structural parameters α, β, σ_1 and σ_2 and with ξ_i as incidental parameters.—If the observable random variables are only partially consistent, then the maximum likelihood estimates of the structural parameters (a) need not be consistent, (b) even if they are consistent and asymptotically normal, alternative estimates may exist which have the same properties but smaller asymptotic variances.—Consistent estimates of structural parameters may be obtained from "modified" equations of maximum likelihood. The lower bound of the variance of estimates of structural parameters, provided by the Cramér-Rao inequality, is attained only on certain conditions which are both necessary and sufficient.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Paul H. Anderson has been appointed Economic Analyst with the Marketing Division, Office of Domestic Commerce, Department of Commerce, Washington.

Dr. Gilbert W. Beebe is now with the Division of Medical Sciences, National Research Council, Washington.

Professor Harald Cramér, Director of the Institute of Mathematical Statistics of the University of Stockholm, was awarded the degree of Doctor of Science, *honoris causa*, by Princeton University on February 22, 1947. Professor Cramér has acted as Visiting Professor of Mathematics at Princeton University and Yale University during the academic year 1946–'47. He will be at the University of California at Berkeley during the 1947 Summer Session.

Dr. Paul M. Densen has accepted a position with the Division of Medical Research Statistics, Bureau of Medicine and Surgery, Veterans Administration, Washington.

Mr. M. V. Divatia is now in charge of the office of the Statistician and Economic Adviser and Under-Secretary to the Government of Sind, Karachi, India.

Mr. Clarence B. Fine, formerly with the Office of Price Administration, has transferred to the Bureau of Old-Age and Survivors Insurance, Social Security Administration, where he is employed as a Sampling Expert.

Prof. Charles C. Grove was appointed Visiting Lecturer in Mathematics at the University of Pennsylvania for the spring semester.

Assoc. Prof. E. E. Haskins of Northeastern University has been appointed to an assistant professorship at the Army Air Forces Institute of Technology, Wright Field, Dayton, Ohio.

Prof. Roger Lessard of the Hull Technical School has accepted a position at the Ecole Polytechnique, Montreal.

Mr. Edward D. Lowery is now a member of the Research Department, Winchester Arms Company, New Haven, Connecticut.

Professor H. B. Mann of Ohio State University has been awarded the Frank Nelson Cole prize in the Theory of Numbers for 1946.

Dr. Margaret P. Martin has been appointed to an assistant professorship in the Department of Preventive Medicine and Public Health, Vanderbilt University Medical School, Nashville, Tennessee.

Dr. A. L. O'Toole is at present employed by the Veterans Administration in the Washington headquarters, as Acting Chief of the Administrative Analysis Division in the Research Service. Dr. O'Toole was released from the Navy on September 23, 1946, to inactive duty in the U. S. Naval Reserve, with the rank

of Commander. Dr. O'Toole served for nearly four years in the Navy, in important administrative and statistical work for the Commander South Pacific Area and South Pacific Force. He will be remembered as having been with Admiral Halsey's Pacific Fleet, and was awarded the Bronze Star Medal. At the time of his release, he was Chief Staff Officer for Commander South Pacific Area and South Pacific Force.

Mr. I. B. Perrott, since his demobilization from the British Army, has been Lecturer in Mathematics at the College of Technology and Commerce, Leicester, England.

Mr. J. S. Ripandelli is now with the Actuarial Department of the Jefferson Standard Life Insurance Company of Greensboro, North Carolina.

Dr. Ronald W. Shephard of the University of California has been appointed to the staff of the Department of Mathematics, New York University.

Mr. John R. Stehn is now a member of the Research Laboratory of the General Electric Company, Schenectady, New York.

Dr. Charles W. Vickery, formerly of Ohio State University, is engaged in work as a Research Consultant in New York City.

———————◆———————

Miss Margaret Jeannin Dix, of the University of California Statistical Laboratory, died an accidental death at her home in Berkeley on June 20, 1946.

Mr. Albert M. Freeman, of the Boston Fiduciary and Research Association, died May 20, 1946.

Dr. Walter Schilling, of the Stanford University Hospital, died suddenly in San Francisco, December 16, 1946.

———————◆———————

Summer Statistical Session at the University of California at Berkeley

The important advances in the theory of statistics during the war and especially the unprecedented growth in the fields of application have created a strong demand for trained statisticians to fill both the research and the teaching positions all over the country. Since in many cases the war time education had to be somewhat sketchy, unsystematic, and not very conducive to a thorough coverage of the vast material, it is felt that a relatively brief set of courses on a rather advanced level would be beneficial to many persons, both those who already hold research or teaching positions in statistics, as well as those who prepare for higher degrees.

With this object in mind, the University of California at Berkeley is offering a set of statistical courses during the Summer Session, June 23rd to August 2nd, 1947. There will be three courses: (i) General Theory of Random Variables and Frequency Distributions, by Harald Cramér of the University of Stockholm;

(ii) Problems of Testing Hypotheses and of Estimation, by J. Neyman, University of California, Berkeley; and (iii) Seminar Course. The last will be given by seven scholars, each giving two hours of lectures, as follows:

1. Statistical Astronomy.	R. J. TRUMPLER
2 Orthogonal Polynomials and Problems of Moments.	G. SZEGÖ
3. Methods of Calculation.	V. F. LENZEN
(a) Gibbs' Methods in Statistical Mechanics	
(b) Darwin-Fowler Method of Statistics.	
4 Large Scale Sampling Surveys.	P. C. MAHALANOBIS
5. Statistical Problems Arising in Nuclear Physics Measurements	R. SERBER
6. Problems of Population Genetics.	S. EMERSON
7. Interactions between Industrial Problems and Mathematical Statistics.	H. SCHEFFÉ

The purpose of the Seminar Course is to introduce the students either to branches of pure mathematics contingent on mathematical statistics but not ordinarily taught in the universities or to various fields of knowledge offering fruitful fields for statistical studies.

Summer Statistical Session at Virginia Polytechnic Institute

A Summer Statistical Session will be held at Virginia Polytechnic Institute, Blacksburg, Virginia, August 5 to September 5, 1947. This Session will be sponsored jointly by Virginia Polytechnic Institute, University of North Carolina, University of Michigan, Iowa State College, and the Federal Bureau of Agricultural Economics.

The faculty will consist of; Walter A. Hendricks, B.A.E., U.S.D.A.; Renis Likert, University of Michigan; H. L. Lucas, University of North Carolina; Maurice G. Kendall, England; George W. Snedecor, Iowa State College; Frank Yates, Rothamsted Experiment Station, England; Earl E. Houseman, B.A.E., U.S.D.A.; Raymond J. Jessen, Iowa State College, and Boyd Harshbarger, Virginia Polytechnic Institute.

The following courses will be offered for credit: Engineering Statistics; Statistical Methods; Design of Animal Experiments; Schedule Design and Interview Techniques for Sample Surveys; Sampling Design and Analysis; Mathematical Theory of Sampling; Seminar; Mathematical Statistics, and Experimental Design.

In addition to the faculty, probable Seminar speakers are: W. F. Callendar, W. G. Cochran, Miss Gertrude M. Cox, W. E. Deming, George Gallup, M. H. Hansen, Harold Hotelling, Arnold King, and Charles F. Sarle.

Inquiries regarding the Summer Session should be addressed to Boyd Harshbarger, Professor of Statistics, Summer Statistical Session, Virginia Polytechnic Institute, Blacksburg, Virginia.

New Members

The following persons have been elected to membership in the Institute
(January 1 to February 28, 1947):

Asofsky, Samuel, B.S. (C.C.N.Y.) Stat., National Jewish Welfare Board, *1256 E. 13 St., Brooklyn 30, N. Y.*

Auer, Richard M., A.M. (Columbia) Instr. in Math., State Teachers Coll., Montclair, N. J., *88 No. 16 St., East Orange*

Bakan, David, M.A. (Indiana) Chief Stat., Comm. on Selection and Training of Aircraft Pilots, National Research Council, *259 Natatorium, Ohio State Univ., Columbus 10, Ohio*

Beatty, Glenn H., A.B. (Ohio State) Grad. student and Fellow, Iowa State College, *Station A, General Delivery, Ames, Iowa*

Campbell, Wallace A., B.S. (Columbia) Stat .Analyst, War Assets Administration, *483 Washington Ave., Brooklyn 16, N. Y.*

Cella, Francis R., M.A. (Kentucky) Assoc. Prof. of Statistics and Director, Bur. of Business Research, Univ. of Oklahoma, Norman, Okla.

Chapman, Douglas G., M.A. (Toronto) Asst. Prof. of Math., Univ. of British Columbia, Vancouver, Canada

Cheydleur, Benjamin F., B.A. (Wisconsin) Chief, Mechanized Analysis, Naval Ordnance Lab., *602 Avenue E, District Heights, Washington 19, D. C.*

Coombs, Clyde H., Ph.D. (Chicago) Ass't Prof. of Psychology, and Research Psychologist, Institute for Human Adjustment, Univ. of Michigan, Ann Arbor, Mich., *1027 E. Huron*

Corton, Edward L., Jr., M.B.A. (Chicago) Grad. student, Iowa State Coll., *803 Hodge Ave., Ames, Iowa*

Davis, Harold., A.B. (Brooklyn Coll.) Stat., Navy Dept., *416—33 St., S.E., Washington, D. C.*

Dutton, Arthur M., B.S.E.E. (Iowa State) Grad. Fellow, Mathematics Dept., Iowa State Coll., Ames, Iowa

Fay, Edward A., A.M. (Harvard) Grad. student, Univ. of California, Berkeley, *415 South 17th St., Apt. 2B, Richmond, Calif.*

Flanagan, John C., Ph.D. (Harvard) Prof. of Psychology, Univ. of Pittsburgh, Pittsburgh 13, Pa.

Gardner, Eric F., Ed.M. (Boston Teachers) Teaching Fellow and Milton Fellow, Grad. School of Educ., Harvard Univ., Cambridge, Mass., *Walker House, 40 Quincy St.*

Gerende, Lincoln J., C.Ph.M., U. S. Navy, *Naval Medical Res. Institute, National Naval Medical Center, Bethesda 14, Md.*

Grossman, Evelyn, M.A. (Columbia) Stat., U. S. Dept. of Agriculture, *6401—14 St., N. W., Washington 12, D. C.*

Hill, Edwin A., Jr., M.A. (Columbia) Instr. in Math., Coll. of the City of N. Y., *50 West 67 St., New York 23, N. Y.*

Horton, H. Burke, M.B.A. (Texas) Senior Transport Analyst, *2906 Naylor Rd., S. E., Washington 20, D. C.*

Horvitz, Daniel G., B.S. (Mass. State) Grad. student, Iowa State Coll., *2137 Country Club Blvd., Ames, Iowa*

Ikhtiar-ul-Mulk, S. M., M.A. (Punjab, India) Grad. student, Princeton Univ., *Graduate College, Princeton, N. J.*

Jaeger, Carol M., B.A. (Dubuque) Statistician, *1300 Columbia Terrace, Peoria 5, Ill.*

Jessen, Raymond J., Ph.D. (Iowa State) Res. Assoc. Prof., Iowa State College, and Agric. Statistician, U.S.D.A., *Statistical Lab., Iowa State Coll., Ames, Iowa*

Kinzer, Mrs. Lydia Greene, M.A. (Kansas) Ass't Instr. in Math., Ohio State Univ., *585 East Town Street, Columbus 15, Ohio*

Langenhop, Carl E., M.S. (Iowa State) Instr. in Math., Iowa State Coll., *Apt. 3, Cranford Annex, Ames, Iowa*

Lowy, Melitta E., A.B. (Hunter) Statistician, Grad. student, Columbia Univ., *645 West End Ave., New York 25, N. Y.*

Mattila, Sakari, Fil.Mag. (Helsinki) High School of Commerce, Helsinki, Finland

Mayerson, Allen L., B.S. (Michigan) Grad. student and Teaching Fellow, Univ. of Mich., *1302 Packard St., Ann Arbor, Mich.*

McCreary, Garnet E., M.A. (Queen's Univ.) Research Fellow, Statistical Lab., Iowa State Coll., Ames, Iowa

McMillan, Olan T., M.A. (Michigan) Instr. in Math., Michigan State Coll., East Lansing, Mich.

Morris, Edward B., A.B. (Indiana) Statistician, U. S. Bur. of Labor Statistics, *1915 Ridge Place S. E., Washington 20, D. C.*

Moshman, Jack, B.A. (New York) Tutor in Math., Queens Coll., Flushing, N. Y., *125–09 Liberty Ave., Richmond Hill 19*

Natrella, Mrs. Mary G., B.A. (Pennsylvania) Statistician, Bureau of Ships, Navy Dept., *1210—12th St., N. W. Washington 5, D. C.*

Neal, T. Ellison, A.B. (Geo. Washington) Statistician, Textile Dev. Dept., U. S. Rubber Co., Hogansville, Ga.

Noble, Carl E., Ph.D. (Iowa) Quality Methods Engineer, Kimberly Clark Corp., Lakeview Mill, Neenah, Wis.

Ostle, Bernard, M.A. (British Columbia) Teaching Ass't, School of Bus. Adm., Univ. of Minnesota, Minneapolis, Minn.

Oxtoby, Toby E., B.A. (Iowa) Grad. Ass't, Dept. of Psychology, State Univ. of Iowa, Iowa City, Iowa

Peisakoff, Melvin P., Student, Princeton Univ., *34 North West College, Princeton, N. J.*

Rothschild, Colette, (Ecole Normale Superieure) Attachee de Recherches au Centre National de la Recherche Scientifique, *43 rue Madame, Paris VI*, France*

Slonim, Morris J., M.B.A. (Harvard) Statistician, Bureau of Labor Statistics, *810 Wayne Place S. E., Washington 20, D. C.*

Soler, Reuben I., B.B.A. (C.C.N.Y.) Statistician, Food and Drug Administration, *246 Portland St., S. E., Washington, D. C.*

Stouffer, Samuel A., Ph.D. (Chicago) Prof. of Sociology and Director of the Laboratory of Social Relations, Emerson Hall, Harvard Univ., Cambridge, Mass.

Teicher, Henry, B.A. (Iowa) Graduate student, Columbia Univ., *139 Osborne Terrace, Newark, N. J.*

Tiedeman, David V., M.A. (Rochester) Instr. in Educ., Grad. School of Educ., Harvard Univ., *Walker House, 40 Quincy St., Cambridge 38, Mass.*

Tintner, Gerhard, Ph.D. (Vienna) Prof. of Economics and Mathematics, Iowa State Coll., Ames, Iowa

Weiss, Eleanor S., Ed.M. (Boston Teachers) Teaching Fellow, Grad. School of Educ., Harvard Univ., *2005 Commonwealth Ave., Brighton 35, Mass.*

Wilson, William A., Jr., A.B. (California) Teaching Ass't in Psychology, Univ. of Calif., Berkeley 4, Calif.

Woodell, Allan D., A.B. (N. Y. State Teachers, Albany) Graduate student in math., Univ. of Mich., *425 Church St., Ann Arbor, Mich.*

Omitted from 1946 lists of new members:

Feraud, Prof. Lucien, Faculte des Sciences Economiques et Sociales, Univ. de Geneve, *24 rue Henri Mussard, Genève, Switzerland*

REPORT ON THE ATLANTIC CITY MEETING OF THE INSTITUTE

The Ninth Annual Meeting of the Institute of Mathematical Statistics was held at Atlantic City, New Jersey, on Friday and Saturday, January 24–25, 1947. The meeting was held in conjunction with meetings of the American Economic Association, American Statistical Association, and the Econometric Society. The following 154 members of the Institute attended the meeting:

Beatrice Aitchison, F. L. Alt, R. L. Anderson, T. W. Anderson, K. J. Arrow, Max Astrachan, B. M. Bennett, Joseph Berkson, A. J. Berman, C. I. Bliss, Paul Boschan, A. E. Brandt, M. F. Bresnahan, Philip Brown, O. P. Bruno, R. W. Burgess, O. K. Buros, B. H. Camp, F. R. Cella, Uttam Chand, K. L. Chung, C. W. Churchman, P. C. Clifford, W. J. Cobb, W. G. Cochran, F. G. Cornell, D. R. Cowan, Harald Cramèr, J. H. Curtiss, J. F. Daly, G. B. Dantzig, D. G. Deihl, D. B. DeLury, B. W. Dempsey, H. F. Dorn, F. W. Dresch, A. J. Duncan, David Durand, P. S. Dwyer, Churchill Eisenhart, W. D. Evans, Will Feller, C. D. Ferris, Irving Fisher, L. R. Frankel, M. A. Geisler, Leon Gilford, M. A. Girshick, C. H. Graves, K. E. Greene, S. W. Greenhouse, F. E. Grubbs, E. T. Gumbel, Margaret Gurney, Louis Guttman, Trygve Haavelmo, K. W. Halbert, M. H. Hansen, Miriam S. Harold, T. E. Harris, Boyd Harshbarger, Bernard Hecht, Wassily Hoeffding, H. B. Horton, Harold Hotelling, E. E. Houseman, Helen M. Humes, Leonid Hurwicz, Seymour Jablon, R. W. James, R. J. Jessen, H. L. Jones, Alice S. Kaitz, H. B. Kaitz, L. S. Kellogg, H. S. Konijn, Tjalling Koopmans, C. F. Kossack, R. L. Kozelka, D. H. Leavens, Howard Levene, J. E. Lieberman, Rensis Likert, S. B. Littauer, Irving Lorge, P. J. McCarthy, P. W. McGann, F. E. McIntyre, H. F. MacNeish, J. D. Maddrill, Jacob Marschak, Max Millikan, A. M. Mood, Mrs. Margaret Moore, J. W. Morse, J. E. Morton, Frederick Mosteller, D. N. Nanda, P. M. Neurath, Jerzy Neyman, M. L. Norden, Nilan Norris, H. W. Norton, P. S. Olmstead, E. G. Olds, Sophie Rakesky, Chester Rapkin, Olav Reiersol, W. A. Reynolds, P. R. Rider, C. F. Roos, A. C. Rosander, Ernest Rubin, Herman Rubin, P. J. Rulon, Frank Saidel, Marion M. Sandomire, Max Sasuly, F. E. Satterthwaite, E. D. Schell, E. M. Schrock, D. H. Schwartz, G. R. Seth, L. W. Shaw, W. A. Shewhart, J. H. Smith, R. T. Smith, Leslie E. Simon, Milton Sobel, C. M. Stein, G. T. Steinberg, Joseph Steinberg, H. W. Steinhaus, F. F. Stephan, A. P. Stergion, M. S. Stevens, G. J. Stigler, S. A. Stouffer, Zenon Szatrowski, B. J. Tepping, J. W. Tukey, D. F. Votaw, Jr., Helen M. Walker, J. H. Watkins, Louis Weiner, Samuel Weiss, S. S. Wilks, Elizabeth W. Wilson, C. P. Winsor, J. Wolfowitz, M. A. Woodbury, Holbrook Working, C. A. Wright, and T. O. Yntema.

The first session, a joint session with the Econometric Society and the Biometrics Section of the American Statistical Association, was held at two o'clock on Friday afternoon, and was devoted to the topic, *Applications of Statistical Techniques to Agricultural Economics.* Holbrook Working of Stanford University presided. The following four papers were presented:

1. *Use of Variance Components in the Analysis of Market Differentials in Hog Prices.* R. L. Anderson, University of North Carolina.
2. *An Application of the Analysis of Variance in the Economic Evaluation of Production.* Boyd Harshbarger, Virginia Polytechnic Institute.
3. *A Model of the Economic Interdependence between Agriculture and the National Economy.* Trygve Haavelmo, Cowles Commission for Research in Economics.
4. *The Reduced-Form Method for Estimating Simultaneous Economic Relationships.* M. A. Girschick, Bureau of the Census.

The session concluded with a discussion of these papers by T. W. Anderson, Columbia University; Milton Friedman, University of Chicago; and, Harold Hotelling, University of North Carolina.

At 8 o'clock on Friday evening there was a joint session with the Econometric Society and the American Statistical Association, on the topic, *When is the Analysis of Variance Useful in Economic Research?* Arthur R. Tebbutt of Northwestern University presided, and the following three papers were presented:

1. *The Advantages of the Analysis of Variance for Research and Managerial Control Purposes.* Harry Pelle Hartkemeier, University of Missouri.
2. *Estimation of Economic Relationships and Multivariate Regression.* Leonid Hurwicz, Iowa State College.
3. *Nonstandard Forms of Variance Analysis.* W. Allen Wallis, University of Chicago.

There was discussion of these papers by Tjalling Koopmans, Cowles Commission for Research in Economics: Gerhard Tintner, Iowa State College; and, J. W. Tukey, Princeton University.

At 10 o'clock on Saturday morning there was a joint session with the American Statistical Association devoted to the topic, *Use of Ordered Observations in Statistical Analysis*, with Harold Hotelling of the University of North Carolina as chairman. The following two papers were presented:

1. *Estimation of Parameters by Use of Order Statistics.* Frederick Mosteller, Harvard University.
2. *Tolerance Limits.* Jacob Wolfowitz, Columbia University.

There was discussion of these papers by John H. Smith, Bureau of Labor Statistics; Howard L. Jones, Illinois Bell Telephone Company; and J. W. Tukey, Princeton University.

At the Saturday morning session one contributed paper of the Institute of Mathematical Statistics was also presented, by E. J. Gumbel, Newark College of Engineering, on the topic: *The Asymptotic Distribution of the Range.*

The Institute's session at 2 o'clock Saturday afternoon was devoted to contributed papers. W. G. Cochran, president of the Institute, presided, and the following four papers were presented:

1. *A Test of Significance of the Coefficient of Rank Correlation for More than Thirty Ranked Items.* Nilan Norris, Hunter College.
2. *A Generalized T Measure of Multivariate Dispersion.* Harold Hotelling, University of North Carolina.
3. *Asymptotic Properties of Maximum and Quasi-Maximum Likelihood Estimates.* Herman Rubin, Cowles Commission for Research in Economics.
4. *The Corner Test for Association.* J. W. Tukey, Princeton University, and Paul Olmstead, Bell Telephone Laboratories.

Abstracts of these papers appear elsewhere in this issue.

Following the session on contributed papers, Professor Jerzy Neyman of the University of California gave an invited address on the topic: *On Consistent Estimates, with Particular Reference to Structural Relations between Several Variables all Subject to Random Error.* A discussion of this address followed, by Miss E. L. Scott, University of California; A. Wald, Columbia University; and Tjalling Koopmans, Cowles Commission for Research in Economics.

The meeting closed with the annual business meeting of the Institute, which was held at 5 p.m. on Saturday in Haddon Hall. Reports by the President, Secretary-Treasurer, and Editor were followed by the election of officers for 1947: Will Feller, President; Morris H. Hansen and John H. Curtiss, Vice-Presidents; and Paul S. Dwyer, Secretary-Treasurer.

P. S. DWYER,
Secretary.

ON THE ASYMPTOTIC DISTRIBUTION OF DIFFERENTIABLE STATISTICAL FUNCTIONS

By R. v. Mises

Harvard University

TABLE OF CONTENTS

Introduction. If n real variables x_1, x_2, $\cdots$, x_n are subject to a probability distribution with the element $dV_1(x_1)dV_2(x_2) \cdots dV_n(x_n)$ one can ask for the distribution of any function f of x_1, x_2, $\cdots x_n$. We are primarily interested in *statistical functions*, i.e. in functions that depend on the *repartition*[1] $S_n(x)$ of the n quantities x_1, x_2, $\cdots x_n$ only. The simplest case is that of the *linear statistical functions*

$$(1) \qquad f = \int \psi(x)\, dS_n(x) = \frac{1}{n}\,[\psi(x_1) + \psi(x_2) + \cdots + \psi(x_n)].$$

The so-called Central Limit Theorem of Probability Calculus states that the distribution of a linear statistical function, if n tends to infinity, approaches more and more the normal (Gauss) distribution if some very general conditions linking $\psi(x)$ and the $V_\nu(x)$ are fulfilled. It has been shown, ten years ago, [2] that the restriction to linear functions here is immaterial. Much more general

---

[1] The function $S_n(x)$ is called the repartition of the real quantities x_1, x_2, $\cdots$, x_n if $nS_n(x)$ is the number of those among the x_1, x_2, $\cdots$, x_n that are smaller than or equal to x.

statistical functions tend towards normalcy with increasing n, for example the variance of mth order

$$(2) \qquad f = M_m = \int (x - a)^m \, dS_n(x), \qquad a = \int x \, dS_n(x)$$

and, likewise, such combinations as the Lexis quotient $M_2/a(1 - a/N)$ or Gini's disparity measure $1 - \int (1 - S_n)^2 dx/a$ or, in the multidimensional case, the correlation coefficient, etc. On the other hand, statistical functions are known whose distributions assume, asymptotically, a form different from the Gaussian. One example is Pearson's Chi-square, another the test function ω^2, introduced by H. Cramér [1] and the author [4]:

$$(3) \qquad f = \omega^2 = \int g'(x)[S_n(x) - \bar{V}_n(x)]^2 \, dx$$

where $g'(x) > 0$ and

$$(4) \qquad \bar{V}_n(x) = \frac{1}{n} [V_1(x) + V_2(x) + \cdots + V_n(x)].$$

N. V. Smirnoff [7, 8] computed the asymptotic distribution of ω^2 for the case that all $V_\nu(x)$ and, therefore, $\bar{V}_n(x)$ equal one and the same distribution function $V(x)$. The result differs widely from the Gaussian distribution.

In order to understand all this it is necessary to consider f as a function defined *in the space of distributions* $V(x)$ (or in a sub-space of it). Then, the variable f whose distribution is sought is the value of $f\{V(x)\}$ at the "point" $S_n(x)$ and should be written as $f\{S_n(x)\}$. Such "functions of functions" were first introduced by Vito Volterra (1887) and are today a familiar topic of higher analysis. The first statement that can be made is that the asymptotic distribution of $f\{S_n(x)\}$ depends mainly on the behavior of $f\{V(x)\}$ at the point $\bar{V}_n(x)$ defined by (4).

Volterra also introduced the notion of derivatives and of Taylor development for a "fonction de ligne." Using these concepts a more specific statement can be pronounced: *The type of asymptotic distribution of a differentiable statistical function $f\{S_n(x)\}$ depends on which is the first non-vanishing term in the Taylor development of $f\{V(x)\}$ at the point $\bar{V}_n(x)$; if it is the linear term the limiting distribution is normal, under restrictions that can easily be derived from the Central Limit Theorem; in other cases higher types of asymptotic distributions result.*

The present paper tries to establish this theorem and to furnish preliminary information about the asymptotic distribution of the *second type*.

If both the function $f\{V(x)\}$ and the sequence of distributions $V_1(x)$, $V_2(x)$, $V_3(x)$, $\cdots$ are defined independently of each other, it cannot be presumed that the derivative of f vanishes at $\bar{V}_n(x)$. In this sense the normal distribution appears as the "general case" of an asymptotic distribution while the higher types represent certain "singularities." In the case of type m, $(m = 1, 2, 3, \cdots)$,

the distribution of the expression

$$(5) \qquad n^{m/2}[f\{S_n(x)\} - f\{\bar{V}_n(x)\}]$$

tends towards a function of bounded mean value and variance. For $m = 1$ it is a Gauss function with mean value 0 and finite variance. For any uneven m the distribution is symmetrical with respect to the zero point. If f is given, the limiting distribution *is essentially determined if in addition to $\bar{V}_n(x)$ one function of two variables, $\bar{U}_n(x, y)$, is known,*

$$(6) \qquad \begin{aligned} \bar{U}_n(x, y) &= \frac{1}{n} \sum_{\nu=1}^{n} [V_\nu(x) - V_\nu(x) V_\nu(y)], & (x \leqq y) \\ &= \frac{1}{n} \sum_{\nu=1}^{n} [V_\nu(y) - V_\nu(x) V_\nu(y)], & (x \geqq y). \end{aligned}$$

For instance, in the case of the linear function $(m = 1)$ defined in eq. (1), the (second order) variance of (5) is found as the Stieltjes integral

$$(7) \qquad \int \psi(x)\psi(y) \, d\bar{U}_n(x, y)$$

and no mean values of higher order are required for computing the moments of any order, whatever m is.

For $m = 2$ the complete expression for the characteristic function of the asymptotic distribution of (5) is developed in Part III of this paper. It has the form

$$(8) \qquad \frac{1}{D(ui)}$$

where $D(\lambda)$ is in general the Fredholm determinant of a symmetrical kernel that depends on the second derivative of $f\{V(x)\}$ at $V = \bar{V}_n$, on $\bar{V}_n$ and on $\bar{U}_n$. If the $V_\nu(x)$ are discontinuous distributions with saltus at k distinct points only, D is the determinant of a quadratic form of k variables. This happens to be the case with Pearson's χ^2 while the ω^2 distribution found by Smirnoff represents a fairly general case of the asymptotic distribution of second type.

PART I. PRELIMINARY THEOREMS

1. Asymptotically equal distributions. Let $K_1, K_2, K_3, \cdots$ be an infinite sequence of collectives, k_n the number of variables in K_n and A_n, B_n two functions of these variables, $(n = 1, 2, 3, \cdots)$. The cumulative distribution functions of A_n and B_n will be denoted by $P_n(x)$ and $Q_n(x)$ respectively, i.e.

$$(1) \qquad P_n(x) = \text{Prob } \{A_n \leqq x\}, \qquad Q_n(x) = \text{Prob } \{B_n \leqq x\}$$

and the expectation of $| A_n - B_n |$ by

$$(2) \qquad E_n\{ | A_n - B_n | \}$$

all these quantities being taken with respect to the distribution in K_n.

Two functions $F_n(x)$ and $G_n(x)$ both depending on the parameter n are said to be *asymptotically equal* if

(3) $$\lim_{n \to \infty} |F_n(x) - G_n(x)| = 0 \quad \text{uniformly in } x.$$

If this is the case for the cumulative distribution functions $P_n(x)$ and $Q_n(x)$ of A_n and B_n we shall also say that A_n and B_n have the same *asymptotic distribution*. Eq. (3) will also be written as $F_n(x) \sim G_n(x)$. The following can be proved:

LEMMA A. *If with increasing n the expectation of the absolute difference between A_n and B_n tends towards zero and if one of the functions $P_n(x)$ or $Q_n(x)$ is asymptotically equal to a function $F_n(x)$ that has a uniformly bounded derivative,* i.e.

(4) $$\lim E_n\{|A_n - B_n|\} = 0, \qquad \frac{dF_n(x)}{dx} < M \quad \text{for all } n$$

then A_n and B_n have the same asymptotic distribution.

This statement, in a slightly different wording, was proved in an earlier paper [2] and the proof will not be repeated here. If one of the various definitions for "stochastical convergence" is used, one can also say that A_n and B_n, under the stated conditions, converge stochastically towards each other.

The Lemma A can be extended and modified in various ways. First, it is obvious that the expectation of $|A_n - B_n|$ can be replaced by that of any positive power $|A_n - B_n|^k$. With respect to F_n one could ask for the existence of a bounded derivative in all points except for a zero set only. Then P_n and Q_n would still converge everywhere except for this zero set and the definition of asymptotically equal distributions could be extended to this case. In the present paper this will not be done as it is not our purpose to strive for results of the possibly greatest generality.

2. Special class of statistical functions: quantics. Preliminary to the study of general statistical functions a special class which corresponds to quantics (homogeneous polynomials) of mth order must be discussed. Let $V_1(x)$, $V_2(x)$, $V_3(x)$, $\cdots$ be the cumulative distribution functions in a sequence of one-dimensional collectives C_1, C_2, C_3, $\cdots$ and $S_n(x)$ the repartition of a sample drawn from the n-dimensional collective K_n, with the distribution element

$$dV_1(x_1)dV_2(x_2) \cdots dV_n(x_n).$$

We introduce

(5) $$T_n(x) = S_n(x) - \bar{V}_n(x), \qquad \bar{V}_n(x) = \frac{1}{n} \sum_{\nu=1}^{n} V_\nu(x).$$

Here, $nT_n(x)$ is obviously the *excess of observed values* $\leq x$ over their expected number. Quantics of first, second, third, $\cdots$ order are then defined as

$$f_1\{S_n(x)\} = \int \psi(x)\, dT_n(x)$$

(6) $$f_2\{S_n(x)\} = \int\int \psi(x, y)dT_n(x)dT_n(y)$$

$$f_3\{S_n(x)\} = \int\int\int \psi(x, y, z)\, dT_n(x)\, dT_n(y)\, dT_n(z)$$

all integrals to be extended over the total range of x. Of course, only such ψ for which the respective integral exists are admitted. The first, f_1, is obviously a linear statistical function and the asymptotic distribution of $\sqrt{n}\, f_1$ is, under well-known conditions, a Gauss function with the mean value zero and the variance given in eq. (7) of the Introduction. In f_2, f_3, $\cdots$ the ψ may be supposed to be symmetrical with respect to their variables. It will be seen later (Part II, sec. 2) that the first derivative of f_2, the first and second derivatives of f_3, etc. vanish at the point $\bar{V}_n(x)$.

All the above functions f_1, f_2, f_3, $\cdots$ can be considered (if the ψ are continuous) as the limits of ordinary quantics in k variables. Choose k disjoint intervals I_1, I_2, $\cdots$, I_k on the x-axis, and call I_{k+1} their complement. Denote the increment of $V_\nu(x)$ within I_κ by $p_{\nu\kappa}$ and the increment of $S_n(x)$ by $\rho_{n\kappa}$. Obviously $p_{\nu\kappa}$ is the probability, within C_ν, of x falling in the interval I_k and $n\rho_{n\kappa}$ is the number of observed sample values in the same interval. We introduce the excess values ξ_λ:

(7) $$\xi_\kappa = \rho_{n\kappa} - \bar{p}_{n\kappa}, \qquad \bar{p}_{n\kappa} = \frac{1}{n}\sum_{\nu=1}^{n} p_{\nu\kappa}$$

and form the sums

(8) $$f_1 = \sum_{\kappa=1}^{k} \psi_\kappa \xi_\kappa, \qquad f_2 = \sum_{\kappa,\lambda}^{1\cdots k} \psi_{\kappa\lambda} \xi_\kappa \xi_\lambda, \qquad f_3 = \sum_{\kappa,\lambda,\mu}^{1\cdots k} \psi_{\kappa\lambda\mu} \xi_\kappa \xi_\lambda \xi_\mu, \cdots .$$

By selecting suitable sets of intervals I_1, I_2, $\cdots$, I_k and appropriate values for the constants ψ_κ, $\psi_{\kappa\lambda}$, $\cdots$, one can approximate the integrals (6) by sums of the form (8).

Our next task will be to find *asymptotic values for the expectation and for the moments of the quantities* defined in (8). Clearly a formula for the expectation of a power product $\xi_1^\alpha \xi_2^\beta \xi_3^\gamma \cdots$ where α, β, γ, $\cdots$ are positive integers, is the only thing we need. To arrive at such a formula we replace each of the one-dimensional collectives C_ν by a k-dimensional C_ν^* in the following way.

In C_ν^* the chance variable is a k-dimensional vector which can take $(k+1)$ distinct values only: it can be zero or coincide with the unit vector parallel to

one of the k axes. To the latter values of the variable we assign the probabilities $p_{\nu 1}$, $p_{\nu 2}$, $\cdots$, $p_{\nu k}$ and to the zero the probability

$$(9) \qquad\qquad p_{\nu,k+1} = 1 - p_{\nu 1} - p_{\nu 2} - \cdots - p_{\nu k}$$

This quantity, of course, may vanish. The mean value of C_ν^* is the point with the coordinates $p_{\nu 1}$, $p_{\nu 2}$, $\cdots$, $p_{\nu k}$.

If the n collectives C_1^*, C_2^*, $\cdots$, C_n^* are combined, the *sum* of the n observed vector values is a vector with the components $n\rho_{n1}$, $n\rho_{n2}$, $\cdots$, $n\rho_{nk}$. If in each C_ν^* the origin is shifted to the mean value and the coordinates with respect to the new origin are called z_1, z_2, $\cdots$, z_k, the sums of the observed z_1, z_2, $\cdots$, z_k-values will be $n\xi_1$, $n\xi_2$, $\cdots$, $n\xi_k$ rather than $n\rho_{n1}$, $n\rho_{n2}$, $\cdots$, $n\rho_{nk}$. Thus it is seen that all questions concerning the distributions of ξ_1, ξ_2, ξ_3, $\cdots$ can be answered on the basis of the well-known rules on the *addition of n independent chance variables*. This leads to the symbolic formula for the expectation:

$$(10) \qquad E_n\{ (n\xi_1)^\alpha (n\xi_2)^\beta (n\xi_3)^\gamma \cdots \} = \left(\sum_{\nu=1}^n Z_{\nu 1} \right)^\alpha \left(\sum_{\nu=1}^n Z_{\nu 2} \right)^\beta \left(\sum_{\nu=1}^n Z_{\nu 3} \right)^\gamma \cdots ,$$

where on the right-hand side each term

$$(11) \qquad\qquad Z_{\nu 1}^\kappa Z_{\nu 2}^\lambda Z_{\nu 3}^\mu \cdots$$

has to be replaced by

$$(11') \qquad\qquad \int z_1^\kappa z_2^\lambda z_3^\mu \cdots \, dV_\nu^*(z).$$

Here, obviously, $V_\nu^*(z)$ is the distribution function in C_ν^* and the expressions $(11')$ are in fact sums of $(k+1)$ terms, for example

$$
\begin{aligned}
(12) \qquad \int z_1 z_2 \, dV_\nu^*(z) &= p_{\nu 1}(1 - p_{\nu 1})(-p_{\nu 2}) + p_{\nu 2}(-p_{\nu 1})(1 - p_{\nu 2}) \\
&\quad + \sum_{\iota=3}^{\lambda+1} p_{\nu\iota}(-p_{\nu 1})(-p_{\nu 2}) = -p_{\nu 1} p_{\nu 2} .
\end{aligned}
$$

It will be seen in the next section that only very few of these sums are needed for computing the asymptotic value of (10). Note that the value of $(11')$ can be expressed in terms of $p_{\nu 1}$, $p_{\nu 2}$, $p_{\nu 3}$, $\cdots$ alone if ξ_1, ξ_2, ξ_3, $\cdots$ only appear in the product.

3. Asymptotic expectation of excess-power products. We first consider the case where the sum of exponents α, β, γ, $\cdots$ is an *even* number

$$(13) \qquad\qquad \alpha + \beta + \gamma + \cdots = 2m.$$

On the right-hand side of (10) stands a sum of n^{2m} terms, each a product of $2m$ factors $Z_{\nu\kappa}$. It follows from $(11')$ that the absolute value of a product cannot surpass 1. The second subscripts are the same in each term: first α ones, then

β twos, γ threes, etc. The first subscripts are in each term a combination of $2m$ digits out of $\nu = 1, 2, 3, \cdots, n$. The number of those combinations which include s different ν-values, $(s = 1, 2, \cdots 2m)$, is

$$(14) \qquad \binom{n}{s} K_s^{(m)} = \binom{n}{s}\left[s^{2m} - \binom{s}{1}(s-1)^{2m} + \cdots + \binom{s}{s-1} 1^{2m} \right].$$

Obviously, the $K_s^{(m)}$ are bounded (independent of n).

If $s > m$ the combination of first subscripts must include at least one ν-value that appears only once. All those products vanish since

$$(15) \qquad \int z_\kappa \, dV_\nu^*(z_\kappa) = 0 \ \text{ for all } \kappa, \nu$$

due to the fact that the origin in the z-space coincides with the mean value of the distribution $V_\nu^*(z)$. Note that

$$(16) \qquad \begin{aligned} \lim_{n \to \infty} \left[\binom{n}{s} : n^m \right] &= 0 & (s < m) \\ &= \frac{1}{m!} & (s = m). \end{aligned}$$

It follows that the sum of all terms in (10) that correspond to any $s < m$ are of the order $o(n^m)$ or smaller.

Thus, we arrive at an asymptotic expression for E_n by dividing both sides of (10) by n^m:

$$(17) \qquad n^m E_n\{\xi_1^\alpha \xi_2^\beta \xi_3^\gamma \cdots\} \sim \frac{1}{n^m} \sum \left(\prod Z_{\nu\kappa} \right)$$

where only such products on the right-hand side are retained which include *exactly m different ν-values each appearing twice.*

In analogy to (12) we compute

$$(18) \qquad \begin{aligned} \int z_\iota z_\kappa dV_\nu^*(z) &= -p_{\nu\iota} p_{\nu\kappa} & (\iota \neq \kappa) \\ &= p_{\nu\iota}(1 - p_{\nu\iota}) & (\iota = \kappa) \end{aligned}$$

and write, for the sake of abbreviation

$$(19) \qquad P_{\iota\kappa}^{(\nu)} = p_{\nu\iota} \delta_{\iota\kappa} - p_{\nu\iota} p_{\nu\kappa} = P_{\iota\kappa}^{(\nu)}$$

with the usual meaning of $\delta_{\iota\kappa}$ ($= 0$ if $\iota \neq \kappa$ and $= 1$ if $\iota = \kappa$). Then the sum to the right in (17) includes $(2m!)/2^m$ terms, each a product of m factors $P_{\iota\kappa}^{(\nu)}$. If each of the m couples ι, κ consists of two different figures, the respective product appears $\alpha! \, \beta! \, \gamma! \cdots$ times; if r couples are doubles ($\iota = \kappa$) the multiplicity of the term is $2^{-r} \alpha! \, \beta! \, \gamma! \cdots$. Therefore, (17) takes the form

$$(20) \qquad n^m E_n\{\xi_1^\alpha \xi_2^\beta \xi_3^\gamma \cdots\} \sim \frac{\alpha! \beta! \gamma! \cdots}{n^m} \sum 2^{-r} P_{\iota_1\kappa_1}^{(\nu_1)} P_{\iota_2\kappa_2}^{(\nu_2)} \cdots P_{\iota_m\kappa_m}^{(\nu_m)}.$$

In this sum the upper indices are any set of m digits out of $1, 2, 3, \cdots, n$ and the subscripts are all sets of m couples including α ones, β twos, γ threes, etc. To each such set of m couples belong $\binom{n}{m}$ terms of the sum. The number of sets of couples is bounded (independent of n). The exponent r is the number of doubles ($\iota = \kappa$) among the m pairs.

The expression (20) admits of a transformation which renders it much more suitable. Assume that a set of couples ι, κ has been chosen according to the conditions and consider the product

$$(21) \qquad \left(\sum_{\nu=1}^{n} P_{\iota_1\kappa_1}^{(\nu)}\right)\left(\sum_{\nu=1}^{n} P_{\iota_2\kappa_2}^{(\nu)} \cdots\right)\left(\sum_{\nu=1}^{n} P_{\iota_m\kappa_m}^{(\nu)}\right)2^{-r}.$$

Among the n^m terms which we obtain by developing (21) are all terms appearing in the sum (20), each of them repeated $m!$ times and, in addition,

$$(22) \qquad n^m - \binom{n}{m}m! = n^m - n(n-1)(n-2) \cdots (n-m+1)$$

other products of m factors P. Since the difference (22) divided by n^m goes to zero with increasing n and each $|P|$ is smaller than 1, the additional terms have no importance. We therefore introduce the quantities

$$(23) \qquad \overline{P}_{\iota\kappa} = \frac{1}{n}\sum_{\nu=1}^{n} P_{\iota\kappa}^{(\nu)} = \delta_{\iota\kappa} - \sum_{\nu=1}^{n} p_{\nu\iota} - \frac{1}{n}\sum_{\nu=1}^{n} p_{\nu\iota} p_{\nu\kappa}.$$

Then (20) can be written as

$$(24) \qquad n^m E_n\{\xi_1^\alpha \xi_2^\beta \xi_3^\gamma \cdots\} \sim \frac{\alpha!\beta!\gamma! \cdots}{m!} \sum_{\iota,\kappa} 2^{-r} \overline{P}_{\iota_1\kappa_1} \overline{P}_{\iota_2\kappa_2} \cdots \overline{P}_{\iota_m\kappa_m}.$$

Here we have a sum of a finite number of terms. It will be supposed in all that follows that *the $\overline{P}_{\iota\kappa}$ as defined in (23) do not vanish identically as n increases indefinitely.*

Since in the sum (24) no upper indices appear, equal terms repeat themselves. We can, therefore, rearrange it, using the polynomial coefficients and absorbing at the same time the factor 2^{-r}. The final form of (24) is given in the following Lemma B_1, which also includes a statement for the case of an *uneven* sum of exponents $\alpha + \beta + \gamma + \cdots$. In fact, it is easily seen that if again half the sum is called m, no group of terms on the right-hand side of (10) exists that would supply a finite limit when divided by n^m. Thus we arrive at

LEMMA B_1. *If $n\xi_\kappa$ is the numerical excess of observed over expected quantities falling in the interval I_κ, the asymptotic expectation of the excess-power product $\xi_1^\alpha \xi_2^\beta \xi_4^\gamma \cdots$ is given by*

$$(\sqrt{n})^{\alpha+\beta+\gamma+\cdots} E_n\{\xi_1^\alpha \xi_2^\beta \xi_3^\gamma \cdots\} \sim 0 \qquad \textit{if } \alpha + \beta + \gamma + \cdots \textit{ uneven}$$

$$(25) \qquad \sim \sum_\sigma \frac{\alpha!\beta!\gamma! \cdots}{\sigma_{11}!\sigma_{22}! \cdots \sigma_{12}! \cdots} (\tfrac{1}{2}\overline{P}_{11})^{\sigma_{11}}(\tfrac{1}{2}\overline{P}_{22})^{\sigma_{22}} \cdots \overline{P}_{12}{}^{\sigma_{12}} \overline{P}_{13}{}^{\sigma_{13}} \cdots,$$

$$\textit{if } \alpha + \beta + \gamma + \cdots \textit{ even}$$

the sum to be extended over all sets of non-negative integers $\sigma_{11}, \sigma_{22}, \cdots, \sigma_{12}, \cdots$
that fulfill the conditions

(25′) $\sigma_{11} = \frac{1}{2}(\alpha - \sigma_{12} - \sigma_{13} - \cdots), \qquad \sigma_{22} = \frac{1}{2}(\beta - \sigma_{21} - \sigma_{23} - \cdots), \cdots$

The $\bar{P}_{\iota\kappa}$ *as defined in* (23) *depend on two groups of mean values only, namely on*

(25″) $\bar{p}_\kappa = \frac{1}{n} \sum_{\nu=1}^{n} p_{\nu\kappa}$ *and* $\overline{p_\iota\, p_\kappa} = \frac{1}{n} \sum_{\nu=1}^{n} p_{\nu\iota}\, p_{\nu\kappa}.$

Some properties of the matrix $\bar{P}_{\iota\kappa}$ will be discussed in the next Section.

For practical computation, instead of (25), a *recursion formula* may be used which follows immediately from (24). Writing simply $(\alpha, \beta, \gamma, \cdots)$ for the sum in (24) the formula reads

(26) $(\alpha, \beta, \gamma, \cdots) = \frac{1}{2}(\alpha - 2, \beta, \gamma, \cdots)\bar{P}_{11} + \frac{1}{2}(\alpha, \beta - 2, \gamma, \cdots)\bar{P}_{22} + \cdots$
$\qquad + (\alpha - 1, \beta - 1, \gamma, \cdots)\bar{P}_{12} + (\alpha, \beta - 1, \gamma - 1), \cdots)\bar{P}_{23} + \cdots.$

If all the original distributions $V_\nu(x)$ are equal, this recursion formula, and from it (25), can be derived almost immediately from the theorem on the multiplication of characteristic functions with the addition of chance variables.

Note that the expectation of the product $\xi_\iota\xi_k$ is $\bar{P}_{\iota\kappa}/n$ for any value of n.

4. Asymptotic expectation and variance of quantics. We first state a characteristic property of the expression (25) for the expectation of an excess power product. Let us denote by $C_{\alpha,\beta,\gamma,\dots}$ the right-hand side of (25) in the case of even $\alpha + \beta + \gamma + \cdots$. Then, if $C_{\alpha,\beta,\gamma,\dots}$ is expressed in terms of $\bar{P}_{\iota\kappa}$ and each time the subscript 2 is changed into 1, we arrive at the value of $C_{\alpha+\beta,0,\gamma,\dots}$. This would not be the case if $C_{\alpha,\beta,\gamma,\dots}$ were expressed in terms of p_ι, since e.g.

$$C_{11} = \bar{P}_{11} = \overline{p_1} - \overline{p_1 p_1}, \qquad C_{12} = \bar{P}_{12} = -\overline{p_1 p_2}.$$

In order to prove the statement we observe that the $C_{\alpha,\beta,\gamma,\dots}$ can be derived from the coefficients in the development of the mth power of a quadric:

(27) $(\frac{1}{2}\sum_{\iota,\kappa} \bar{P}_{\iota\kappa}\, t_\iota\, t_\kappa)^m = m! \sum \dfrac{C_{\sigma,\beta,\gamma,\dots}}{\alpha!\beta!\gamma! \dots}\, t_1^\alpha\, t_2^\beta\, t_3^\gamma \cdots$

It follows that

(27′) $C_{\alpha,\beta,\gamma,\dots} = \dfrac{1}{m!} \dfrac{\partial^{2m}}{\partial t_1^\alpha\, \partial t_2^\beta\, \partial t_3^\gamma \cdots} [(\frac{1}{2}\sum_{\iota,\kappa} \bar{P}_{\iota\kappa}\, t_\iota\, t_\kappa)^m].$

If in the subscripts of $\bar{P}_{\iota\kappa}$ the ones and twos are identified, the quadric becomes a function of $t_1 + t_2, t_3, t_4, \cdots$ and the derivative with respect to $\partial t_1^\alpha\, \partial t_2^\beta$ equals the derivative with respect to $\partial t_1^{\alpha+\beta}$. On the other hand, the latter derivative corresponds to the value of $C_{\alpha+\beta,0,\gamma,\dots}$ in the form (27′).

Taking $m = 2$, $\alpha = \beta = \gamma = \delta = 1$, eq. (25) supplies

(28) $n^2 E_n\{\xi_\iota\xi_\kappa\xi_\lambda\xi_\mu\} \sim \bar{P}_{\iota\kappa}\bar{P}_{\lambda\mu} + \bar{P}_{\iota\lambda}\bar{P}_{\kappa\mu} + \bar{P}_{\iota\mu}\bar{P}_{\kappa\lambda} \dots$

According to the above statement this is correct whether ι, κ, λ, μ are or are not different from each other. Thus, if $\psi_{\iota\kappa\lambda\mu}$ is a symmetric set of constants, we have

$$(28') \qquad n^2 E_n \Big\{ \sum_{\iota,\cdots\kappa} \psi_{\iota\kappa\lambda\mu} \xi_\iota \xi_\kappa \xi_\lambda \xi_\mu \Big\} \sim 3 \sum_{\iota\cdots\mu} \psi_{\iota\kappa\lambda\mu} \overline{P}_{\iota\kappa} \overline{P}_{\lambda\mu}.$$

In general, the numerical factor to the right, i.e. the number of sets of couples drawn from $2m$ figures, is $(2m)!/2^m m! = 1\cdot3 \cdots (2m-1)$. Thus we can state:

LEMMA B_2. *If a quantic f_{2m} is defined according to* (8) *with symmetric coefficients, its asymptotic expectation is given by*

$$(29) \qquad n^m E_n\{f_{2m}\} \sim 1.3.5 \cdots (2m-1) \sum \psi_{\iota_1\iota_2\cdots\iota_{2m}} \overline{P}_{\iota_1\iota_2} \overline{P}_{\iota_3\iota_4} \cdots \overline{P}_{\iota_{2m-1}\iota_{2m}}.$$

Before applying this to the continuous case defined in (6), let us consider some characteristic properties of the matrix $\overline{P}_{\iota\kappa}$. According to the definition (19) of $P_{\iota\kappa}^{(\nu)}$ we have

$$(30) \qquad \sum_{\iota,\kappa}^{1\cdots k} P_{\iota\kappa}^{(\nu)} l_\iota l_\kappa = \sum_{\iota=1}^{k} p_{\nu\iota} l_\iota^2 - \Big(\sum_{\iota=1}^{k} p_{\nu\iota} l_\iota \Big)^2$$

and using (9) one easily derives from Schwarz' inequality

$$\Sigma p_{\nu\iota} l_\iota^2 - (\Sigma p_{\nu\iota} l_\iota)^2 \geqq p_{\nu,k+1} \Sigma p_{\nu\iota} l_\iota^2.$$

Since $\overline{P}_{\iota\kappa}$ is the arithmetical mean of the $P_{\iota\kappa}^{(\nu)}$ it follows that the matrix $\overline{P}_{\iota}^{\kappa}$ is at least semi-definite and is positive definite except when all $p_{\nu,k+1} = 0\cdot$ In the latter case (if e.g. the k intervals cover the whole x-axis) one has

$$(31) \qquad \sum_{\iota,\kappa}^{1\cdots k} \overline{P}_{\iota\kappa} = \frac{1}{n} \sum_{\nu=1}^{n} \Big[\sum_{\iota=1}^{k} p_{\nu\iota} - \Big(\sum_{\iota=1}^{k} p_{\nu\iota} \Big)^2 \Big] = \frac{1}{n} \sum_{\nu=1}^{r} p_{\nu,k+1}(1 - p_{\nu,k+1}) = 0$$

which shows that here the reciprocal matrix $\overline{P}_{\iota\kappa}^*$ does not exist.

In the "complete" case, that is, with all $p_{\nu,k+1} = 0$, the elements in each horizontal or vertical line of the matrix $\overline{P}_{\iota\kappa}$ have the sum zero. It follows that the k homogenous equations $\Sigma \overline{P}_{\iota\kappa} x_\iota = 0$ have the solution $x_1 = x_2 = \cdots = x_k$ and, therefore, that the cofactors of all elements of $\overline{P}_{\iota\kappa}$ have one and the same value. For each single ν the determinant of $P_{\iota\kappa}^{(\nu)}$ can be computed:

$$|P_{\iota\kappa}^{(\nu)}| = p_{\nu1} p_{\nu2} \cdots p_{\nu k} p_{\nu,k+1}$$

If this is applied to the principal minors of the same determinant in the case $p_{\nu,k+1} = 0$, one finds the characteristic equation of the matrix $P_{\iota\kappa}^{(\nu)}$ to be

$$|\delta_{\iota\kappa} - \lambda P_{\iota\kappa}^{(\nu)}| = -\frac{d}{d\lambda}[(1 - \lambda p_{\nu1})(1 - \lambda p_{\nu2}) \cdots (1 - \lambda p_{\nu k})].$$

This shows that $(k-1)$ characteristic roots separate the abscissas $1/p_{\nu1}$, $1/p_{\nu2}$, $\cdots$, $1/p_{\nu k}$ (one root being zero).

The number k of intervals has nothing to do with the preceding argument leading to the eqs. (25) to (28). Also can the entire computation be repeated

in terms of $dT_n(x_1)$, $dT_n(x_2)$, $dT_n(x_3)$, $\cdots$ instead of ξ_1, ξ_2, ξ_3, $\cdots$ if appropriate differentials are substituted for the $P_{\iota\kappa}$. To find the latter ones we note that $p_{\nu\kappa}$ stands for the increment $dV_\nu(x)$. Thus, using $\delta(x, y)$ in analogy to $\delta_{\iota\kappa}$ ($= 1$ for $x = y$ and $= 0$ for $x \neq y$) we set

$$
(32) \quad
\begin{aligned}
dU_\nu(x, y) &= \delta(x, y)\, dV_\nu(x) - dV_\nu(x)\, dV_\nu(y) \\
&= \delta(x, y)\, dV_\nu(x) - dW_\nu(x, y)
\end{aligned}
$$

which is equivalent to the definition of a function of 2 variables:

$$
(33) \quad
\begin{aligned}
U_\nu(x, y) &= V_\nu(x) - V_\nu(x)V_\nu(y) = V_\nu(x) - W_\nu(x, y) \qquad (x \leq y) \\
&= V_\nu(y) - V_\nu(x)V_\nu(y) = V_\nu(y) - W_\nu(x, y) \qquad (x \geq y).
\end{aligned}
$$

Then $P_{\iota\kappa}$ has to be replaced by

$$
(34) \qquad d\overline{U}_n(x, y) = \frac{1}{n} \sum_{\nu=1}^{n} dU_\nu(x, y) = \delta(x, y)\, d\overline{V}_n(x) - d\overline{W}_n(x, y).
$$

This $d\overline{U}_n(x, y)$ is the expectation of $dT_n(x)\, dT_n(y)/n$.

The function

$$
(35) \qquad\qquad \overline{U}_n(x, y) = \frac{1}{n} \sum_{\nu=1}^{n} U_\nu(x, y)
$$

is the difference of two cumulative distribution functions, one corresponding to a distribution along the straight line $x = y$ with the element $d\overline{V}_n(x)$ and another distribution over the whole plane with the element

$$
(35') \qquad\qquad d\overline{W}_n(x, y) = \frac{1}{n} \sum_{\nu=1}^{n} dV_\nu(x)\, dV_\nu(y).
$$

To each one-dimensional distribution $V_\nu(x)$ belongs one *"distribution excess"* $U_\nu(x, y)$ as defined in (33). The $P_{\iota\kappa}^{(\nu)}$ are the increments of $U_\nu(x, y)$ within the product interval $dxdy$. It is seen from the preceding argument that the asymptotic moments of any quantic (6) or (8) depend only on the average $\overline{U}_n$ of the distribution excesses U_ν.

If a quantic is defined by (6) and the integrals on both sides exist, the asymptotic expectation of f_{2m} may be written in formal analogy to (29) as

$$
(36) \quad
\begin{aligned}
n^m E_n\{f_{2m}\} \sim 1.3.5 \cdots (2m-1) \iint \cdots \int \psi(x_1, x_2, \cdots, x_{2m}) \\
\times\, d\overline{U}_n(x_1, x_2)\, d\overline{U}_n(x_3, x_4) \cdots d\overline{U}_n(x_{2m-1}, x_{2m}).
\end{aligned}
$$

This formula is identical with (29) if ψ has constant values in a finite number of intervals and vanishes outside these intervals. But it will be seen in the next section that (36) can be used in more general cases also.

For the sake of practical computation one may develop the righthand side

of (36) into terms explicitly depending on the given averages $\bar{V}_n(x)$ and $\bar{W}_n(x, y)$. For example, in the case $m = 3$:

$$n^3 E_n\{f_6\} \sim 1.3.5 \iiint [\psi(x_1, x_1, x_2, x_2, x_3, x_3)\, d\bar{V}_n(x_1)\, d\bar{V}_n(x_2)\, d\bar{V}_n(x_3)$$

(37)
$$- 3\psi(x_1, x_1, x_2, x_2, x_3, x_4)\, d\bar{V}_r(x_1)\, d\bar{V}_n(x_2)\, d\bar{W}_n(x_3, x_3)$$

$$+ 3\psi(x_1, x_1, x_2, x_3, x_4, x_5)\, d\bar{V}_n(x_1)\, d\bar{W}_n(x_2, x_3)\, d\bar{W}_n(x_4, x_5)$$

$$- \psi(x_1, x_2, x_3, x_4, x_5, x_6)\, d\bar{W}_n(x_1, x_2)\, d\bar{W}_n(x_3, x_4)\, d\bar{W}_n(x_5, x_6)]$$

In the general case, the numerical factors in the m-tuple integral are the binomial coefficients of order m.

The higher moments of quantics f_m can be computed in the same way as $E_n\{f_m\}$ since any power of f_m is a quantic again. The formulas, however, become more involved since the coefficients of f_m^* are not immediately given in a symmetric form. It will suffice to show here how the (second order) variance of f_2 can be found. The second moment is the expectation of

$$(39) \qquad f_2^2 = \iiiint \psi(x, y)\psi(z, u)\, dT_n(x)\, dT_n(y)\, dT_n(z)\, dT_n(u).$$

Applying here eq. (28) we have

$$n^2 E_n\{f_2^2\} \sim \iint \psi(x, y)\psi(z, u)[\, d\bar{U}_n(x, y)\, d\bar{U}_n(z, u)$$

(40)
$$+ d\bar{U}_n(x, z)\, d\bar{U}_n(y, u) + d\bar{U}_n(x, u)\, d\bar{U}_n(y, z)].$$

The first term in the brackets leads to the square of $n\, E_n\{f_2\}$ while the second and third terms, due to the symmetry of $\Psi(x, y)$, supply two equal integrals. Thus

$$\text{Var }\{nf_2\} \sim 2 \iint \psi(x, y)\psi(z, u)\, d\bar{U}_n(x, z)\, d\bar{U}_n(y, u) =$$

(41)
$$2 \left[\iint \psi(x, x)\psi(y, y)\, d\bar{V}_n(x)\, d\bar{V}_n(y) - 2 \iint \psi(x, y)\psi(y, z)\, d\bar{V}_n(y)\, d\bar{W}_n(x, z) \right.$$

$$\left. + \iint \psi(x, y)\psi(z, u)\, d\bar{W}_n(x, z)\, d\bar{W}_n(y, u) \right].$$

In the same way moments and variances of any order can be computed for any quantic f_m.

5. Final statement on the limit of expectation of quantics. We shall prove the following:

LEMMA B_3 . *Given a sequence of distributions* $V_1(x)$, $V_2(x)$, $V_3(x)$, $\cdots$ *and a quantic of order* $2m$

$$f_{2m} = \iint \cdots \int \psi(x_1 , x_2 , \cdots , x_{2m}) \, dT_n(x_1) \, dT_n(x_2) \cdots dT_n(x_{2m})$$

assume that there exist a continuous function $\Psi(x)$ *and a distribution* $V(x)$ *such that*

(42)
$$| \psi(x_1 , x_2 , \cdots x_{2m}) | \leqq \Psi(x_1) \, \Psi(x_2) \cdots \Psi(x_{2m})$$

$$dV_\nu(x) \leqq dV(x) \text{ for } \quad | x | > X, \quad \nu = 1, 2, 3, \cdots$$

and that the integrals

(42′)
$$\int \Psi^r(x) \, dV(x), \quad (r = 1, 2, \cdots 2m),$$

have finite values. *Then, for any* $\delta > 0$

(43)
$$\lim_{n = \infty} n^{m \, \delta} E_n\{f_{2m}\} = 0.$$

This lemma, on which the main theorem of Part II is based, will be established if it is shown that the formula (36) holds true for functions ψ satisfying the conditions (42).

In the transition from the complete expression (10) for the expectation E_n to the asymptotic value (25) two essential steps were made. First, certain products of the form (11) have been omitted and, second, certain products of $P_{\iota \kappa}^{(\nu)}$ as defined in (19) have been arbitrarily added. This was allowed because each of the products was seen to be smaller than 1 and their number was of the order $O(n^{m-1})$. If a quantic in integral form (6) is considered which involves an infinite number of expressions like (10), a sharper estimate is necessary.

It is easily seen that each integral (11′) is a polynomial in $p_{\nu\kappa}$ including the product $p_{\nu 1}p_{\nu 2}p_{\nu 3} \cdots$ and another factor which is certainly bounded whatever the $p_{\nu\kappa}$ are. Thus, if the expectation of $\xi_1\xi_2 \cdots \xi_{2m}$ is computed, each term of the form (11′) consists of a finite factor and the product $p_{\nu 1}p_{\nu 2} \cdots p_{\nu,2m}$. In passing to the expectation of the quantic, the $p_{\nu\kappa}$ have to be replaced by $dV_\nu(x_\kappa)$ and each neglected term in (10) leads to an expression like

(45)
$$\iint \cdots \int \psi(x_1 , x_2 , \cdots , x_{2m}) \, dV_{\nu_1}(x_1) \, dV_{\nu_2}(x_2) \cdots dV_{\nu_\kappa}(x_\kappa).$$

According to the assumptions of B_3 this integral has a finite value. The number of neglected terms being of the order $O(n^{m-1})$ the omission of these terms is justified.

On the other hand, products of $P_{\iota \kappa}^{(\nu)}$ equal, except for the sign, products of $p_{\nu\iota}p_{\nu\kappa}$ as long as $\iota \neq \kappa$ and, except for a finite factor, products of $p_{\nu\kappa}$ as often as $\iota = \kappa$. Again it is seen that the arbitrarily added terms sum up to integrals

of the form (45). This shows that here too, if the conditions of B_3 are fulfilled, the procedure leading to (25) may be applied.

It follows that, under the conditions (42), if the integral (42') has a finite value, eq. (36) is correct and (43) is an immediate consequence of it. On the other hand, it is obvious that weaker conditions than those given in B_3 would suffice to establish (43).

6. Theorem on products of n functions. The principal source of all explicit formulas on asymptotic distributions lies in certain properties of products of a great number of factors. Laplace devoted a part of his fundamental Treatise of Probability to these problems, but a complete outline of all results from a modern point of view is still lacking. In the third part of the present paper, a rather simple statement on this line will be used which may be formulated here as

LEMMA C. *Let $F_\nu(z_1, z_2, \cdots, z_k)$, $(\nu = 1, 2, 3, \cdots)$, be a sequence of analytic functions of k complex variables and G_n the product $F_1 F_2 \cdots F_n$. Suppose that at the point $z_1 = z_2 = \cdots z_k = 0$ all F_ν have the value 1, vanishing first derivatives, and the second derivatives*

$$(46) \qquad A_{\iota\kappa}^{(\nu)} = \frac{\partial^2 F_\nu}{\partial z_\iota \partial z_\kappa}.$$

Then

$$(47) \qquad \lim_{n \to \infty} \left[G_n\left(\frac{z_1}{\sqrt{n}}, \frac{z_2}{\sqrt{n}}, \cdots, \frac{z_n}{\sqrt{n}}\right) - \exp\left(\frac{1}{2n} \sum_{\iota,\kappa,\nu} A_{\iota\kappa}^{(\nu)} z_\iota z_\kappa\right) \right] = 0$$

uniformly in each bounded region $|z_\iota| \leq Z$ in which the absolute values of the third derivatives of all F_ν have an upper bound M.

In fact, the Taylor development of F_ν supplies under the conditions stated:

$$(48) \qquad F_\nu(z_1, z_2, \cdots, z_k) = 1 + \tfrac{1}{2} \sum_{\iota,\kappa} A_{\iota\kappa}^{(\nu)} z_\iota z_\kappa + O(Z^3)$$

and, therefore,

$$(48') \qquad \log F_\nu(z_1, z_2, \cdots, z_k) = \tfrac{1}{2} \sum_{\iota,\kappa} A_{\iota\kappa}^{(\nu)} z_\iota z_\kappa + O(Z^3).$$

If here all z_ι are replaced by $z_\iota/\sqrt{n}$ and the equations added for $\nu = 1, 2, \cdots, n$ we obtain

$$(49) \qquad \log G_n\left(\frac{z_1}{\sqrt{n}}, \frac{z_2}{\sqrt{n}}, \cdots, \frac{z_k}{\sqrt{n}}\right) = \frac{1}{2n} \sum_{\iota,\kappa,\nu} A_{\iota\kappa}^{(\nu)} z_\iota z_\kappa + nO\left(\frac{Z^3}{n\sqrt{n}}\right).$$

and this shows that the brackets on the left-hand side of (47) are $O(Z/\sqrt{n})$.— It is obvious that (47) would still hold if the condition concerning the third derivatives is replaced by a somewhat weaker one.

PART II. DIFFERENTIABLE STATISTICAL FUNCTIONS

1. Definitions. We consider a one-dimensional cumulative distribution function $V(x)$ as a point in the V-space. If two points $V_1(x)$ and $V_2(x)$ are given the functions

$$(1) \qquad V_1(x) + t[V_2(x) - V_1(x)], \qquad 0 \leq t \leq 1$$

represent the straight segment between $V_1(x)$ and $V_2(x)$. A subset of the V-space that includes all segments determined by its elements is called a *convex* domain.

Now, assume that a sequence of collectives with the distributions $V_1(x)$, $V_2(x)$, $V_3(x)$, $\cdots$ be given. We shall consider functions $f\{V(x)\}$ defined in a convex domain that includes particularly: (1) all average distributions $\bar{V}_n(x)$

$$(2) \qquad \bar{V}_n(x) = \frac{1}{n} \sum_{\nu=1}^{n} V_\nu(x)$$

at least from a certain n on; (2) all repartitions $S_n(x)$ that can occur, i.e. the repartitions of n quantities that belong to the label sets of the given collectives (e.g. positive x, etc.). If $V^0(x)$ and $V(x)$ are any two points of the domain, the quantity

$$(3) \qquad F(t) = f\{V^0(x) + t[V(x) - V^0(x)]\}, \qquad 0 \leq t \leq 1$$

is a function of the real variable t. It will be supposed to admit derivatives with respect to t up to the order $r + 1$.

Following Volterra [9, 10] we define (in a slightly modified way) the derivative f' of a statistical function f in analogy to the set of partial derivatives of a function of several variables. If $V(x)$ would stand for a set of distinct variables $V_1, V_2, V_3, \cdots$ and $V^0(x)$ for their initial values $V_1^0, V_2^0, V_3^0, \cdots$ one would have

$$\frac{d}{dt} f\{V^0(x) + t[V(x) - V(x)]\}_{t=0} = \sum_\nu \frac{\partial f}{\partial V_\nu} (V_\nu - V_\nu^0)$$

where $\partial f/\partial V_\nu$ is the partial derivative of f with respect to V_ν taken at the point $V_\nu = V_\nu^0$. Thus we write

$$(4) \qquad \frac{d}{dt} f\{V^0(x) + t[V(x) - V^0(x)]\}_{t=0} = \int f'\{V^0(x), y\} d(V - V^0)(y)$$

and call f' which depends on $V^0(x)$ and on a scalar variable y, but not on $V(x)$, the (first) *derivative* of $f\{V(x)\}$ at the point $V^0(x)$. Only if a relation (4) is fulfilled for any two points of the convex domain, f is called a (one time) differentiable function.

The derivative of a linear function

$$(5) \qquad A = \int \alpha(x) \, dV(x), \qquad B = \int \beta(x) \, dV(x),$$

is simply the factor $\alpha(y)$, $\beta(y)$ $\cdots$ respectively, independent of the point at which the derivative is taken. If f given as a function of $A, B, \cdots$ one has

(6) $$f'\{V(x), y\} = \frac{\partial f}{\partial A} \alpha(y) + \frac{\partial f}{\partial B} \beta(y) + \cdots.$$

The derivative of the non-linear function

(7) $$f = \int\int \psi(x, y) \, dV(x) \, dV(y)$$

is

(8) $$f'\{V^0(x), y\} = \int [\psi(x, y) + \psi(y, x)] \, dV^0(x).$$

Note that an additive constant in f' (i.e. a quantity independent of y) has no significance since the integral of $d(V - V^0)$ vanishes. It follows from (6) that the first derivative of the mth order variance as defined in (2) of the Introduction, at the point $V^0(x)$ is

(9) $$(y - a_0)^m - my \int (x - a_0)^{m-1} \, dV^0(x)$$

where a_0 is the mean value of $V^0(x)$.

In the same way derivatives of higher order can be introduced. The second derivative of $f\{V(x)\}$ is a function of $V^0(x)$, i.e. of the point at which the derivative is taken, and of two scalar variables y, z which correspond to the two subscripts in the case of a function of distinct variables. The definition of $f''\{V(x), y, z\}$ is given in the equation

(10)
$$\frac{d^2}{dt^2} f\{V^0(x) + t[V(x) - V^0(x)]\}_{t=0}$$
$$= \int\int f''\{V^0(x), y, z\} \, d(V - V^0)(y) \, d(V - V^0)(z).$$

The second derivative of a linear function is zero. The function (7) has the second derivative $\psi(z, y) + \psi(y, z)$ independently of $V^0(x)$. The mth order variance gives, twice differentiated

(11) $$-2mz(y - a_0)^{m-1} + m(m - 1)yz \int (x - a_0)^{m-2} \, dV^0(x).$$

The variables y and z in f'' or in any additive term of f'' may be interchanged and a term depending on one of them may be added or omitted. Thus, f'' can always be written as a symmetric function of y, z without linear terms Accordingly, the second derivative of (7) is also $2\psi(y, z)$.

The derivative of rth order of f at the point $V^0(x)$ will be defined by the equation

(12)
$$\frac{d^r}{dt^r} f\{ V^0(x) + t[V(x) - V^0(x)] \}_{t=0}$$
$$= \iint \cdots \int f^{(r)}\{ V^0(x), y_1, y_2, \cdots, y_r \} \, d(V - V^0)(y_1) \cdots d(V - V^0)(y_r).$$

Here, for given $V^0(x), f^{(r)}$ may be supposed to be a symmetric function of the r variables $y_1, y_2, \cdots, y_r$. The rth derivative of the mth order variance is

(13)
$$\frac{(-1)^r m!}{(m - r + 1)!} y_1 y_2 \cdots y_r$$
$$\times \left[(m - r + 1) \int (x - a_0)^{m-r} \, dV^0(x) - \sum_{\kappa=1}^{r} \frac{(y_\kappa - a_0)^{m-r+1}}{y_\kappa} \right].$$

In the case $r = m$ the expression becomes independent of $V^0(x)$, viz.

(13')
$$(-1)^m m! \, y_1 y_2 \cdots y_m (1 - m)$$

where terms depending on less than r of the variables $y_1, y_2, \cdots, y_r$ have been omitted.

If the definitions (4), (10), (12) are confronted one can see that $f''\{V, y, z\}$ is the first derivative of $f'\{V, y\}$ etc. For proofs see [9] and [10].

2. Taylor development. The function $F(t)$ defined in (3) admits the development

(14) $\quad F(1) - F(0) = F'(0) + \frac{1}{2!} F''(0) + \cdots + \frac{1}{r!} F^{(r)}(0) + \frac{1}{(r+1)!} F^{(r+1)}(\vartheta)$

where ϑ is some quantity between zero and one. According to (3) the left-hand side equals the difference $f\{V(x)\} - f\{V^0(x)\}$. The expressions $F'(0), F''(0), \cdots,$ $F^{(r)}(0)$ are the derivatives as defined in eqs. (4), (10), (12). In the last term to the right, one has to introduce the distribution

(15) $\qquad V'(x) = V^0(x) + \vartheta[V(x) - V^0(x)]$

and then to take the $(r + 1)$st derivative of f at the point $V'(x)$.

For a given $V^0(x)$ each one of the terms on the right-hand side of (14) is a function of $V(x)$. Except for the last one—in which ϑ depends in a certain way on $V(x)$—they are *quantics* with respect to $V(x) - V^0(x)$, of the same kind as those considered in Part I. (There we had S_n instead of V and $\bar{V}_n$ instead of V^0).

The rth term of (14) can be written as

(16) $F_r = \dfrac{1}{r!} \displaystyle\iint \cdots \int \psi(x_1, x_2, \cdots, x_r)\, d(V - V^0)(x_1) \cdots d(V - V^0)(x_r)$

where according to (12)

(16') $\psi(x_1, x_2, \cdots, x_r) = f^{(r)}\{V^0(x), x_1, x_2, \cdots, x_r\}.$

To find the characteristic properties of F_r we compute its derivatives at a point $V_1(x)$. To do this we must replace in (16) the $V(x)$ by

$$V_1(x) + t[V(x) - V_1(x)]$$

then differentiate the product

(17) $\displaystyle\prod_{\kappa=1}^{r} d[(V_1 - V^0)(x_\kappa) + t(V - V_1)(x_\kappa)]$

with respect to t, and finally set $t = 0$. The derivative consists of r terms the first of which will be

$$d(V - V_1)(x_1) \prod_{\kappa=2}^{r} d(V_1 - V^0)(x_\kappa).$$

Due to the fact that ψ may be supposed as a symmetric function, all r terms supply the same integral. Thus the derivative of F_r with respect to t at the point $t = 0$ can be written as

$$\dfrac{1}{(r-1)!} \iint \cdots \int \psi(x_1, x_2, \cdots, x_r)\, d(V - V_1)(x_1) \prod_{\kappa=2}^{r} d(V_1 - V^0)(x_\kappa).$$

Comparing this with the formula (4) which defines the first derivative of a statistical function and writing y instead of x and $V(x)$ instead of $V_1(x)$, we find

(18)

$F_r'\{V(x), y\} =$

$\dfrac{1}{(r-1)!} \displaystyle\iint \cdots \int \psi(y, x_2, x_3, \cdots, x_r)\, d(V - V^0)(x_2) \cdots d(V - V^0)(x_r).$

This is the first derivative of $F_r\{V(x)\}$ at the point $V(x)$. It vanishes at the point $V(x) = V^0(x)$.

The integral in (18) has the same form as that in (14) except that its multiplicity is $(r - 1)$ rather than r. Thus it is immediately seen how the higher derivatives of F_r can be found. For the second derivative $F_r''\{V(x), y, z\}$ we have simply to replace $(r - 1)!$ in (18) by $(r - 2)!$, then x_2 by z and finally to omit in the product the differential $d(V - V^0)(x_2)$. This procedure can be continued up to the derivative of order $(r - 1)$. The rth derivative, finally,

will be

(19) $\qquad F_r^{(r)}\{V(x), y_1, y_2, \cdots, y_r\} = \psi(y_1, y_2, \cdots, y_r)$

independent of $V(x)$ and, according to (16'), equal to the rth derivative of $f\{V(x)\}$ at the point $V^0(x)$. It is also seen that all integrals of the form (16) or (18) vanish if $V(x)$ equals $V^0(x)$. The results can be summarized as follows: The sth term, ($s = 1, 2, \cdots r$), of the development (14) is a function of $V(x)$ for which all derivatives at the point $V^0(x)$ except that of order s vanish while this one equals the sth derivative of the original function $f\{V(x)\}$ at $V^0(x)$. The complete analogy of (14) with the Taylor development of a function of distinct variables is thus evident.

If we assume that $f\{V(x)\}$ is a function whose first $(r - 1)$ derivatives vanish at the point $V^0(x)$, eq. (14) takes the form

$$V(x) - V^0(x) = \frac{1}{r!} \int\int \cdots \int f^{(r)}\{V^0(x), y_1, y_2, \cdots, y_r\}$$

$$\cdot d(V - V^0)(y_1) \cdots d(V - V^0)(y_r)$$

(20)

$$+ \frac{1}{(r+1)!} \int\int \cdots \int f^{(r+1)}\{V'(x), y_1, y_2, \cdots, y_{r+1}\}$$

$$\cdot d(V - V^0)(y_1) \cdots d(V - V^0)(y_{r+1}).$$

By applying to this formula the lemmas A and B of Part I, we shall arrive at the general theorem on asymptotic distributions that is the principal goal of this paper.

3. General theorem. The main result to be derived in the general theory of asymptotic distributions is that the so-called normal distribution represents the first element in an infinite sequence which includes the asymptotic distributions of all differentiable statistical functions, except certain irregular cases. The Gauss distribution covers in fact only those functions whose Taylor development starts with the first (linear) term, in particular the linear statistical functions themselves. If the first $(r - 1)$ terms in the development vanish, the asymptotic distribution of type r becomes valid.

THEOREM I: *Let $V_1(x), V_2(x), V_3(x), \cdots$ be an infinite sequence of distributions and $f\{V(x)\}$ a statistical function with derivatives up to order $(r+1)$. Denote by $S_n(x)$ the repartition of the n label values in the collective with the distribution element $dV_1(x), dV_2(x) \cdots dV_n(x)$ and by $\bar{V}_n(x)$ the arithmetical mean of $V_1(x), V_2(x), \cdots, V_n(x)$. If for large n the first $(r - 1)$ derivatives of $f\{V(x)\}$ at the point $\bar{V}_n(x)$ vanish and the rth derivative equals $\psi_n(y_1, y_2, \cdots, y_r)$, then the distribution of*

(21) $\qquad A_n = n^{r/2}[f\{S_n(x)\} - f\{\bar{V}_n(x)\}]$

is asymptotically equal to the distribution of the rth order quantic

$$B_n = \frac{n^{r'2}}{r!} \int\int \cdots \int \psi_n(x_1, x_2, \cdots, x_r)$$

(22)

$$\cdot d(S_n - \bar{V}_n)(x_1)\, d(S_n - \bar{V}_n)(x_2) \cdots d(S_n - \bar{V}_n)(x_r)$$

under the following conditions:

   a) *The distribution of* (22) *has a uniformly bounded derivative for all* n;

   b) *Within a convex domain in the V-space that includes all* $\bar{V}_n(x)$ *from a certain n on, and all* $S_n(x)$ *that can occur, the* $(r + 1)$*st derivative of* $f\{V(x)\}$ *is smaller in absolute value than a product* $\Psi(y_1)\Psi(y_2) \cdots \Psi(y_{r+1})$ *whereby the integrals* $\int[\Psi(x)]^k\, dV_\nu(x)$ *for* $k = 1, 2, \cdots, 2(r+1)$ *have a finite upper bound for* $\nu = 1, 2, 3, \cdots$.

In order to prove this we introduce in eq. (20) $S_n(x)$ for $V(x)$ and $\bar{V}_n(x)$ for $V^0(x)$, and multiply both sides by $n^{r'2}$. Using the notations (21) and (2) and writing T_n for $(S_n - \bar{V}_n)$, the equation reads

(32)
$$A_n - B_n = \frac{n^{r/2}}{(r+1)!}$$

According to Lemma A the theorem will be verified if we can show that the expectation of the absolute value of the right-hand expression in (23) tends to zero.

According to the Schwarz inequality one has, for any real C:

(24)
$$E_n\{|\,C\,|\} \leqq \sqrt{E_n\{C^2\}}.$$

For fixed values of $\bar{V}_n$ and S_n the integral on the right-hand side of (23) is a quantic of order $(r + 1)$ with the coefficients $\psi_{r+1}(y_1, y_2, \cdots, y_{r+1})$. The square of this integral is a quantic of order $2(r + 1)$ whose coefficients are a finite number (depending only on r) of terms each of which is a product of two ψ_{r+1}-values implying $2(r + 1)$ variables $y_1, y_2, \cdots, y_{2(r+1)}$. The absolute value of these coefficients is, therefore, according to the condition b) smaller than a finite factor times the product $\Psi(y_1)\ \Psi(y_2) \cdots \Psi(y_{2(r+1)})$ and thus fulfills the condition of lemma B_3. If the right-hand side of (23) is identified with C, the expectation of C^2 is, except for a finite factor, the product of n^r times the expectation of the above-mentioned quantic of order $2(r + 1)$. It then follows from lemma B_3 that the limit of $E_n\{C^2\}$ is zero and from (24):

$$\lim_{n \to \infty} E_n\{|\,C_n\,|\} = \lim_{r \to \infty} E_n\{|\,A_n - B_n\,|\} = 0.$$

This accomplishes the proof of Theorem I.

If we apply here what was shown in Part I about the asymptotic distribution of a quantic, we can also state the following.

THEOREM II: *Under the conditions of Theorem I, the asymptotic distribution of a differentiable statistical function $f\{S_n(x)\}$ is essentially determined by*

a) *the average distribution $\bar{V}_n(x)$;*

b) *the first non-vanishing derivative of $f\{V(x)\}$ at the point $\bar{V}_n(x)$;*

c) *the average distribution excess*

(25)
$$\bar{U}_n(x, y) = \bar{V}_n(x) - \frac{1}{n} \sum_{\nu=1}^{n} V_\nu(x) V_\nu(y), \qquad x \leqq y$$

$$= \bar{V}_n(y) - \frac{1}{n} \sum_{\nu=1}^{n} V_\nu(x) V_\nu(y), \qquad x \geqq y.$$

By "essentially determined" is meant determined except for an additional function whose moments of any order are zero. The statement then follows from Theorem I in connection with the fact that the asymptotic moments of quantics have been computed in Part I from the values of $\bar{U}_n(x, y)$.

That functions with all moments vanishing exist has been known for a long time. A simple example given by Shohat and Tamarkin [6] is the following. Let κ be a positive constant smaller than $\frac{1}{2}$, and $u = x^\kappa$, $k = \tan \kappa\pi$. Then, the density (positive or negative)

(26)
$$\varphi(x) = e^{-u} \sin (ku) = Im \, e^{-u(1-ki)}$$

fulfills the condition. In fact, the nth moment of (26) is the (vanishing) imaginary part of the integral

(27)
$$\frac{1}{\kappa} \int_0^\infty u^{(n+1/\kappa)-1} e^{-u(1-ki)} \, du = \frac{(-1)^n}{\kappa} \frac{1}{} (\cos \kappa\pi)^{(n+1/\kappa)} \Gamma\left(\frac{n+1}{\kappa}\right).$$

Since $\varphi(x)$ takes negative values of the amount e^{-u} it can be superimposed to a given distribution density only in cases where the original density remains greater than some multiple of $e^{-u} = \exp(-x^k)$. It can be shown that the moment problem is determinate (i.e. the distribution determined by the moments in a unique way) if the density vanishes at infinity at a sufficiently strong degree.

From the standpoint of statistical theory two distributions with the same moments throughout may be considered as equivalent. This justifies the terminology used in Theorem II. On the other hand, Theorem I is independent of this restriction: The asymptotic distribution of the statistical function $f\{S_n(x)\}$ is under the given conditions identical with that of the corresponding quantic of mth order. A detailed discussion of the case $m = 2$ will be given in Part III. Here follow some illustrations for the general case.

4. Illustrations. The existence of asymptotic distributions of higher types can be exemplified in a comparatively simple way if we start from any known asymptotic distribution of a statistical function.

Let us assume that $g\{V(x)\}$ is a function fulfilling the condition

(28)
$$g\{\bar{V}_n(x)\} = 0$$

for all n, and that the asymptotic c.d.f. for $g\{S_n(x)\}$ is known. There will be some positive integer r such that

$$(29) \qquad \text{Prob } [g\{S_n(x)\} \leqq zn^{-r/2}] \sim \Phi_n(z).$$

If, for instance, g is a linear statistical function r will be 1 and, under well-known conditions, $\Phi_n(x)$ a normal (Gaussian) c.d.f. with finite variance depending on n.

Now, let f be an ordinary function of g and thus another statistical function which may be denoted by $f\{V(x)\}$. According to the rules of differentiation we have

$$(30) \qquad f'\{V(x), y\} = \frac{df}{dg} g'\{V(x), y\}$$

and analogous relations can be derived for the derivatives of higher order. In particular, the following statement, valid in ordinary differential calculus, holds true: If $g\{V(x)\}$ has derivatives of every order and if the first s derivatives of f with respect to g vanish at some point $g = g\{V_1(x)\}$ then also the s first derivatives of f with respect to $V(x)$ will be zero at $V(x) = V_1(x)$. In this way we can devise statistical functions, with vanishing derivatives, for which the asymptotic distribution is known.

For the sake of simplicity we may assume that (29) holds with $r = 1$ and that $f(g)$ is a monotonic increasing function, given in the form

$$(31) \qquad f(g) = g^s[1 + \alpha(g)]$$

with s a positive integer, and the inverse function

$$(31') \qquad g(f) = f^{1/s}[1 + \beta(f)]$$

where $\beta(f)$ goes to zero with $f \to 0$. Then, from (29):

$$(32) \qquad \text{Prob } [f\{S_n(x)\} \leqq zn^{-(s/2)}] \sim \Phi_n(z').$$

if z and z' are connected by

$$n^{-\frac{1}{2}}z' = g(n^{-(s/2)}z) = n^{-\frac{1}{2}}z^{1/s}[1 + \beta(n^{-(s/2)}z)].$$

It follows that

$$z' - z^{1/s} \sim 0$$

and if $\Phi_n(z')$ is supposed to be continuous, (32) becomes

$$(33) \qquad \text{Prob } [f\{S_n(x)\} \leqq zn^{-(s/2)}] \sim \Phi_n(z^{1/s}).$$

This is a distribution of type s.

Take as an example for g the arithmetical mean

$$(34) \qquad g\{S_n(x)\} = \frac{x_1 + x_2 + \cdots + x_n}{n} - \bar{a}_n$$

where x_1, x_2, $\cdots$, x_n are the observed values and $\bar{a}_n$ is the arithmetical mean of the mean values of $V_\nu(x)$. Then, under certain restrictions for the $V_\nu(x)$, there exists a bounded sequence h_n^2 so that

$$\text{Prob}[\sqrt{n}\,\varrho \leq z] \sim \Phi_n(z) = \frac{h_n}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-h_n^2 u^2} \, du.$$

Now if we choose

$$f = 6(\varrho - \sin \varrho) = \varrho^3 \left(1 - \frac{\varrho^2}{20} + \cdots\right)$$

the asymptotic distribution of f will be given by

$$\text{Prob}\,[n\sqrt{n}\,f \leq z] \sim \Phi_n(\sqrt[3]{z}) = \frac{h_n}{\sqrt{\pi}} \int_{-\infty}^{z^{1/3}} e^{-h_n^2 u^2} \, du$$

with the probability density

$$\frac{h_n}{3\sqrt{\pi}}\, z^{-(2/3)}\, e^{-h_n^2 z^{2/3}}.$$

Similar examples can be drawn from the asymptotic distribution of $n\chi^2$ if one asks for the distribution of appropriate functions of $n\chi^2$, etc.

PART III. SECOND-TYPE ASYMPTOTIC DISTRIBUTION

1. Statement of the problem. We now propose to study the asymptotic distribution of a quantic of second order as defined in eq. (6) of Part I. It has been shown in Part II that this covers the case of any statistical function of which the first but not the second derivative at the critical point vanishes.

Independently of what was said before, the problem can be stated in the following way. Given a function $\psi(x, y)$ and a sequence of cumulative distribution functions $V_1(x)$, $V_2(x)$, $V_3(x)$ $\cdots$. Let $\bar{V}_n(x)$ be the arithmetical mean of $V_1(x)$, $V_2(x)$, $\cdots$, $V_n(x)$ and $S_n(x)$ the repartition of a sample z_1, z_2, $\cdots$, z_n drawn from the collective with the distribution element $dV_1(z_1)\, dV_2(z_2)$, $\cdots$, $dV_n(z_n)$, that is: $nS_n(x)$ is the number of those of the observed values $z_1, z_2, \cdots, z_n$ that are smaller than or equal to x. Then the quantity

$$(1) \qquad f = \iint \psi(x, y)\, dT_n(x)\, dT_n(y), \quad \text{where } T_n(x) = S_n(x) - \bar{V}_n(x)$$

is determined by the observations z_1, z_2, $\cdots$, z_n. We ask for the distribution of f at large values of n.

Without loss of generality, the function $\psi(x, y)$ can be supposed to be symmetrical. If, in particular, $\psi(x, y) = \psi(x)\psi(y)$, the quantity f becomes the square of

$$(2) \qquad \int \psi(x)\, dT_n(x) = \frac{1}{n} \sum_{\nu=1}^{n} \left[\psi(z_\nu) - \int \psi(z)\, dV_\nu(x)\right]$$

and its asymptotic distribution can be computed in the manner shown in the last
section of Part I.   Another example would be

$$\psi(x, y) = g(x) \qquad\qquad (x \leq y)$$

(3)

$$= g(y) \qquad\qquad (x \geq y).$$

In this case, integration by parts shows that

(4)
$$f\{S_n(x)\} = \int g'(x) T_n^2(x)\, dx$$

where g' is the derivative of g.   This is the statistical function that takes the
place of χ^2 in continuous problems.   See Introduction eq. (3).

Note that the "excess" $T_n(x)$ vanishes at $x = \pm \infty$ and that for sufficiently
large x the increment $dT_n(x)$ equals $-d\bar{V}_n(x)$.   Thus, conditions for the exist-
ence of the integrals in (1), (2), (4), etc. can be expressed in terms of the given
functions $\psi(x, y)$ and $V_\nu(x)$.

We shall first study the special case that implies so-called discontinuous chance
variables.   In our terminology it is the function $\psi(x, y)$ that has to be specified.
Let $I_1, I_2, \cdots, I_k$ be k mutually exclusive one-dimensional intervals (or groups
of intervals) and I_{k+1} their complement.   Assume that $\psi(x, y)$ has a constant
value when x falls in I_ι and y falls in I_κ, $(\iota, \kappa = 1, 2, \cdots, k+1)$.   The increments
of $S_n(x)$, $\bar{V}_n(x)$, $T_n(x)$ in the interval I_κ will be called ρ_κ, $\bar{p}_\kappa$, ξ_κ respectively.
Clearly, $n\rho_\kappa$ is the number of observed values falling in I_κ, $n\bar{p}_\kappa$ is the expected
number of such values, and $n(\rho_\kappa - \bar{p}_\kappa) = n\xi_\kappa$ the excess of observed over expected
numbers.   Note that the given distributions $V_\nu(x)$ determine increments $p_{\nu\kappa}$
in the interval I_κ and that

(5)
$$\bar{p}_\kappa = \frac{1}{n}(p_{1\kappa} + p_{2\kappa} + \cdots + p_{n\kappa}).$$

Since the sum of all ξ_κ must be zero we can replace ξ_{k+1} by

(6)
$$\xi_{k+1} = -\xi_1 - \xi_2 - \cdots - \xi_k.$$

Thus, the integral (1) can now be written as a sum of k^2 terms

(7)
$$f\{S_n(x)\} = \sum_{\iota,\kappa}^{1\cdots k} \psi_{\iota\kappa}\, \xi_\iota\, \xi_\kappa$$

like that introduced in the second eq. (8) of Part I.

Our next task will be to find the asymptotic distribution of (7) which depends
on the matrix $\psi_{\iota\kappa}$, $(\iota, \kappa = 1, 2, \cdots, k)$, and on the succession of probability
values $p_{\nu\kappa}$, $(\nu = 1, 2, 3, \cdots; \kappa = 1, 2, \cdots k)$.   The matrix $\psi_{\iota\kappa}$ in k variables
will be supposed to be symmetrical.

2. Characteristic function.   We define our chance variable as

(8)
$$x = \frac{n}{2} f.$$

All summations, here and in what follows, are to be extended from 1 to k if not otherwise indicated.   If $P_n(x)$ is the c.d.f. of x, that is

$$(9) \qquad \text{Prob} \left\{ \frac{n}{2} f \leqq x \right\} = P_n(x)$$

the characteristic function (c.f.) is defined by

$$(10) \qquad Q_n(u) = E\{e^{xui}\} = \int e^{xui} dP_n(x).$$

In order to compute Q_n we assume that the quadratic form (8) is transformed, by a linear transformation, into a sum of squares.   Using appropriate (in general complex) coefficients $\alpha_{i\kappa}$ one can write

$$(11) \qquad x = \frac{n}{2} (\eta_1^2 + \eta_2^2 + \cdots + \eta_k^2), \qquad \eta_i = \sum_\kappa \alpha_{i\kappa} \xi_\kappa .$$

(The form $\psi_{i\kappa}$ is here supposed to be non-singular which, however, means no loss of generality).   It will be seen later that explicit knowledge of the $\alpha_{i\kappa}$ is not needed.

Now, for any real or complex y, the identity holds:

$$(12) \qquad e^{\frac{1}{2}y^2} = \frac{1}{\sqrt{2\pi}} \int e^{\frac{1}{2}t^2 + yt} \, dt.$$

If we write v for $\sqrt{ui}$ and replace in (12) successively y by $v\sqrt{n}\,\eta_1$, $v\sqrt{n}\,\eta_2$, $\cdots$ we find

$$(13) \qquad e^{xui} = (2\pi)^{-k/2} \iint \cdots \int \exp\left[-\tfrac{1}{2}\sum t_\kappa^2 + v\sqrt{n}\sum z_\kappa \xi_\kappa\right] dt_1 \, dt_2 \cdots dt_k$$

where

$$(14) \qquad \sum z_\kappa \xi_\kappa = \sum \eta_\kappa t_\kappa, \qquad z_\kappa = \sum_i \alpha_{i\kappa} t_i, \qquad (\kappa = 1, 2, \cdots, k).$$

Since the first exponential factor in the integrand is a constant with respect to the chance variable, the expected value of e^{xui} is given by

$$(15) \qquad Q_n(u) = E\{e^{xui}\} = (2\pi)^{-k/2} \iint \cdots \int \exp\left[-\tfrac{1}{2}\sum t_\kappa^2\right] G_n \, dt_1 \, dt_2 \cdots dt_k$$

with

$$(16) \qquad G_n = E\{ \exp[v\sqrt{n} \sum z_\kappa \xi_\kappa]\}.$$

In order to find G_n we consider the following n collectives C_1 , C_2 , $\cdots$, C_n with discontinuous, $(k+1)$-valued distributions: In C_r the label values are z_1 , z_2 , $\cdots$, z_k , and z_{k+1} , with $z_{k+1} = 0$, their probabilities p_{r1} , p_{r2} , $\cdots$, $p_{r,k+1}$. The c.f. of this distribution at the point $-iv/\sqrt{n}$ is

$$(17) \qquad \sum_{\kappa=1}^{k+1} p_{r\kappa} e^{vz_\kappa/\sqrt{n}}.$$

If we multiply the n expressions (17) for $\nu = 1, 2, \cdots n$ the product will be—according to well-known rules of probability calculus—the c.f. for the distribution of the *sum* of the n label components in the collective formed by combining $C_1, C_2, \cdots, C_n$. This sum is

$$\sum n \rho_\kappa z_\kappa$$

and therefore,

$$(18) \qquad E\left\{\exp\left[\frac{v}{\sqrt{n}} \sum n\rho_\kappa z_\kappa\right]\right\} = \prod_{\nu=1}^{n}\left[\sum_{\kappa=1}^{k+1} p_{\nu\kappa} e^{vz_\kappa/\sqrt{n}}\right].$$

Multiplying both sides of this equation by

$$(19) \qquad \exp\left[-\frac{v}{\sqrt{n}} \sum n\bar{p}_\kappa z_\kappa\right] = \exp\left[-\sum_{\nu=1}^{n}\frac{v}{\sqrt{n}} \sum_\kappa p_{\nu\kappa} z_\kappa\right]$$

and using the abbreviation

$$(20) \qquad \bar{z}_\nu = \sum_\kappa p_{\nu\kappa} z_\kappa$$

we arrive at

$$(21) \qquad G_n = E\{\exp[v\sqrt{n}\sum \xi_\kappa z_\kappa]\} = F_1 F_2 \cdots F_n$$

with

$$(22) \qquad F_\nu = \sum_{\kappa=1}^{k+1} p_{\nu\kappa} e^{v(z_\kappa - \bar{z}_\nu)/\sqrt{n}}.$$

This solves the problem: By inserting (21), (22) in (15) and carrying out the integration with respect to $t_1, t_2, \cdots, t_k$ one has expressed $Q_n(u)$ in terms of the given $p_{\nu\kappa}$ and of the coefficients $\alpha_{\iota\kappa}$ which link the z_λ to the t_κ. This expression for $Q_n(u)$ holds for all n.

We have still to show that the integral (15) exists, at least for small $|u|$ or $|v|$, independently of the value of n. For this purpose we develop F_ν, as given in (22), in the neighborhood of $v = 0$. At this point $F_\nu = 1$ and the first derivative vanishes by virtue of (20). We thus have

$$(23) \qquad F_\nu = 1 + \frac{v^2}{2n} \sum_{\kappa=1}^{k+1} p_{\nu\kappa}(z_\kappa - \bar{z}_\nu)^2 e^{\vartheta_\kappa v(z_\kappa - \bar{z}_\nu)/\sqrt{n}}$$

with $|\vartheta_\kappa| \leqq 1$. From the definition of z_κ in (14) it follows that the ratio $|z_\kappa|/T$ with

$$T^2 = t_1^2 + t_2^2 + \cdots + t_k^2$$

has an upper bound depending on the $\alpha_{\iota\kappa}$ only. On the other hand, according to (20), $\bar{z}_\nu$ is a weighted mean of the z_κ and, therefore, $|z_\kappa - \bar{z}_\nu|$ will not surpass twice the maximum $|z_\kappa|$:

$$(25) \qquad |z_\kappa - \bar{z}_\nu| < \alpha T$$

where α is a positive function of the coefficients $\alpha_{\iota\kappa}$ which, in turn, are determined by the $\psi_{\iota\kappa}$. Introducing (25) in (23) we find

$$| F_\nu | < 1 + \frac{| v |^2 \alpha^2 T^2}{2n} e^{|v|\alpha T/\sqrt{n}} \leqq e^{|v^2|\alpha^2 T^2/n}$$

and, finally, from (21):

$$(26) \qquad\qquad | G_n | < e^{|v|^2 \alpha^2 T^2} = e^{|u| \alpha^2 T^2}.$$

Thus it is seen that for

$$(27) \qquad\qquad | u | < \frac{1}{2\alpha^2} \quad \text{or} \quad 1 - 2\alpha^2 | u | \geqq \eta^2 > 0$$

the integral (15) admits the upper bound

$$(28) \qquad | Q_n(u) | < (2\pi)^{-k/2} \iint \cdots \int e^{-\eta^2 T^2/2} \, dt_1, dt_2, dt_k = \eta^{-k}.$$

It also follows that the contribution to $Q_n(u)$ from the region $T > T_0$ tends to zero with increasing T_0, uniformly with respect to n and with respect to u in the region $| u | < 1/2\alpha^2$.

3. Asymptotic value of $Q_n(u)$. If the quantity F_ν introduced in (22) is considered as a function of $z_1/\sqrt{n}$, $z_2/\sqrt{n}$, $\cdots$, $z_k/\sqrt{n}$, we may write

$$(29) \qquad\qquad F_\nu(z_1, z_2, \cdots, z_k) = \sum_{\kappa=1}^{k+1} p_{\nu\kappa} e^{v(z_\kappa - \bar z_\nu)}.$$

Here, $\bar z_\nu$ is defined by (20) and, on the right-hand side, z_{k+1} is zero. These functions $F_\nu(z_1, z_2, \cdots, z_k)$ for $\nu = 1, 2, 3, \cdots$ have all the properties required in Lemma C of Part I: At the point $z_1 = z_2 = \cdots = z_k = 0$ one has $F_\nu = 1$, the first derivatives are

$$\frac{\partial F_\nu}{\partial z_\iota} = v p_{\nu\iota} - v p_{\nu\iota} \sum_{\kappa=1}^{k+1} p_{\nu\kappa} = 0$$

and the second derivatives, ($\iota \neq \kappa$),

$$(30) \qquad \begin{aligned} \frac{\partial^2 F_\nu}{\partial z_\iota^2} &= v^2 p_{\nu\iota}(1 - p_{\nu\iota}) - v^2 p_{\nu\iota}\left[p_{\nu\iota} - p_{\nu\iota} \sum_{\kappa=1}^{k+1} p_{\nu\kappa} \right] = v^2 p_{\nu\iota}(1 - p_{\nu\iota}) \\ \frac{\partial^2 F_\nu}{\partial z_\iota \partial z_\kappa} &= v^2 p_{\nu\iota}(-p_{\nu\kappa}) - v^2 p_{\nu\iota}\left[p_{\nu\kappa} - p_{\nu\kappa} \sum_{\lambda=1}^{k+1} p_{\nu\lambda} \right] = -v^2 p_{\nu\iota} p_{\nu\kappa}. \end{aligned}$$

The third derivatives are certainly bounded in any finite region of the z-space, and this means also in any finite region of the t-space.

The matrix of the second derivatives except for the factor v^2 is exactly that defined in eq. (19) of Part I:

$$(31) \qquad\qquad P_{\iota\kappa}^{(\nu)} = p_{\nu\iota}\delta_{\iota\kappa} - p_{\nu\iota}p_{\nu\kappa}$$

and the arithmetical means of the derivatives from the matrix in eq. (23) of Part I:

$$(31') \qquad \bar{P}_{\iota\kappa} = \frac{1}{n} \sum_{\nu=1}^{n} p_{\nu\iota}\,\delta_{\iota\kappa} - \frac{1}{n} \sum_{\nu=1}^{n} p_{\nu\iota}\,p_{\nu\kappa}.$$

Applying Lemma C we find

$$(32) \qquad G_n = G_n\left(\frac{z_1}{\sqrt{n}}, \frac{z_2}{\sqrt{n}}, \cdots, \frac{z_k}{\sqrt{n}}\right) \sim \exp\left[\frac{v^2}{2} \sum_{\iota,\kappa} \bar{P}_{\iota\kappa} z_\iota z_\kappa\right].$$

This is valid in any finite t-region. Since it has been shown at the end of the foregoing section that, for small $|v|$, the outside contribution to the integral (15) converges uniformly (for all n) towards zero, we are allowed to introduce (32) in (15). Writing

$$(33) \qquad \sum_{\iota,\kappa} \bar{P}_{\iota\kappa} z_\iota z_\kappa = \sum_{\iota,\kappa} \gamma_{\iota\kappa} t_\iota t_\kappa, \qquad \text{whereby } \gamma_{\iota\kappa} = \sum_{\lambda,\mu} \bar{P}_{\lambda\mu}\, \alpha_{\iota\lambda}\, \alpha_{\kappa\mu}$$

equation (15) becomes

$$(34) \quad Q_n(u) \sim (2\pi)^{-k/2} \iint \cdots \int \exp\left[-\tfrac{1}{2}\sum_{\kappa} t_\kappa^2 + \tfrac{1}{2}ui \sum_{\iota,\kappa} \gamma_{\iota\kappa} t_\iota t_\kappa\right] dt_1\, dt_2 \cdots dt_k.$$

Now, it is well known that if $m_{\iota\kappa}$ is any positive definite matrix with the determinant $|m_{\iota\kappa}|$, then

$$(35) \quad (2\pi)^{-k/2} \iint \cdots \int \exp\left[-\tfrac{1}{2}\sum_{\iota,\kappa} m_{\iota\kappa} t_\iota t_\kappa\right] dt_1\, dt_2 \cdots dt_k = \frac{1}{\sqrt{|m_{\iota\kappa}|}}.$$

This is likewise true if the matrix $m_{\iota\kappa}$, which we also call M, has the form $M = M_1 - \lambda M_2$ where M_1 is positive definite, M_2 arbitrary (complex) and $|\lambda|$ sufficiently small. Thus, the integration formula (35) applies to (34) and the result is reached, for small $|u|$:

$$(36) \qquad Q_n(u) \sim Q(u) = \frac{1}{\sqrt{D(ui)}} \quad \text{with} \quad D(\lambda) = |\delta_{\iota\kappa} - \lambda\gamma_{\iota\kappa}|.$$

If the $\alpha_{\iota\kappa}$ which transform the given quadric into a sum of squares are known, (36) with (33) supply the solution of our problem.

The formula (36) is susceptible of several useful transformations. Let us write A for the matrix $\alpha_{\iota\kappa}$, A' for the transposed matrix, and Ψ, $\bar{P}$, Γ, I respectively for the matrices $\psi_{\iota\kappa}$, $\bar{P}_{\iota\kappa}$, $\gamma_{\iota\kappa}$, $c_{\iota\kappa}$. Then, obviously

$$(37) \qquad \Psi = A'A, \qquad \Gamma = A\bar{P}A', \qquad M = I - ui\Gamma.$$

If we multiply M by A' to the left and by A to the right, we obtain

$$(38) \qquad A'MA = A'IA - ui\,A'A\bar{P}A'A = \Psi - ui\,\Psi\bar{P}\Psi.$$

In this operation the determinant of M is multiplied by $|\psi_{\iota\kappa}|$. Thus $D(\lambda)$ can be written as

$$(39) \qquad D(\lambda) = \frac{|\psi_{\iota\kappa} - \lambda\gamma'_{\iota\kappa}|}{|\psi_{\iota\kappa}|} \quad \text{with} \quad \gamma'_{\iota\kappa} = \sum_{\lambda,\mu} \psi_{\iota\lambda} \bar{P}_{\lambda\mu} \psi_{\mu\kappa}.$$

Here, the knowledge of the $\alpha_{\iota\kappa}$ is no longer required.

If the matrix (38) is multiplied twice by Ψ^*, the inverse of Ψ, we find $\Psi^* - ui\bar{P}$ and, therefore,

$$(40) \qquad D(\lambda) = |\psi_{\iota\kappa}| \times |\psi_{\iota\kappa}^* - \lambda\bar{P}_{\iota\kappa}|.$$

As $\bar{P}$ is positive definite and Ψ^* real, it follows that all roots of $D(\lambda)$—the "Eigenwerte" of Γ—are real numbers. Therefore, $D^{-1/2}(ui)$ is a regular function along the real axis in the u-plane. Thus, (36) which was proved so far for small $|u|$ only remains valid for all real values of u: The c.f. of the asymptotic distribution is represented by $D^{-1/2}(ui)$ for all real u-values.

Multiplying (38) only once by Ψ^* we obtain one of the two forms

$$(41) \qquad I - ui\ \Psi\bar{P} \quad \text{or} \quad I - ui\ \bar{P}\Psi$$

which lead to

$$(42) \qquad D(\lambda) = |\delta_{\iota\kappa} - \lambda s_{\iota\kappa}| = |\delta_{\iota\kappa} - \lambda s_{\kappa\iota}|, \qquad s_{\iota\kappa} = \sum_{\mu} \psi_{\iota\mu} \bar{P}_{\mu\kappa}.$$

Although this formula has been derived by means of Ψ^* it can be seen by continuity considerations that it remains valid whatever the (symmetric) matrix $\psi_{\iota\kappa}$ is. The formula makes it clear that the asymptotic distribution of the quadric $\Sigma\psi_{\iota\kappa}\xi_\iota\xi_\kappa$ is completely determined by the "Eigenwerte" of the matrix $S = \Psi\bar{P}$. This bears out our second main theorem in Chapter II, as far as quartics of the form (8) are concerned. It will be seen in sec. 5 how (42) applies to the continuous case.

We, finally, apply to (36) a transformation that is valid only if $\bar{P}$ has an inverse matrix $\bar{P}^*$. (As shown in Part I, sec. 4 this is not the case if the k intervals to which the subscripts 1, 2, $\cdots$, k refer cover the whole range of the variables x_1, x_2, $\cdots$, x_n). Multiplying (41) by $\bar{P}^*$ we find the matrix $\bar{P}^* - ui\Psi$ and thus

$$(43) \qquad D(\lambda) = |\bar{P}_{\iota\kappa}| \times |\bar{P}_{\iota k}^* - \lambda\psi_{\iota\kappa}|.$$

This is equivalent to

$$(44) \qquad Q(u) = |\bar{P}_{\iota\kappa}|^{1/2} \int\int \cdots \int \exp\left[-\tfrac{1}{2}\Sigma\bar{P}_{\iota\kappa}^*\xi_\iota\xi_\kappa \right.$$
$$\left. + \tfrac{1}{2}ui\sum\psi_{\iota\kappa}\xi_\iota\xi_\kappa\right] d\xi_1 d\xi_2 \cdots d\xi_k.$$

According to the definition of the characteristic function eq. (44) can be interpreted as stating that

$$(45) \qquad |\bar{P}_{\iota\kappa}|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\Sigma\bar{P}_{\iota\kappa}^* \xi_\iota\xi_\kappa\right]$$

is the asymptotic probability density for the simultaneous occurrence of ξ_1, ξ_2, $\cdots$, ξ_k. The expression (45) can be arrived at by applying the Central Limit Theorem to the case of k independent chance variables. Since, however, $\bar{P}^*$ does not exist in general, eq. (44) would not be a suitable point of departure for developing the theory that concerns us here.

4. Asymptotic value of $P_n(x)$, illustrations. The relationship between the c.f. and the c.d.f. of a distribution is well known and need not be discussed here in detail. We shall use, in this section, two aspects of this relationship only. First, the continuity theorem, first proved by G. Pólya [5], stating that if the c.f. $Q_n(u)$ tend towards a limiting function $Q(u)$, the corresponding c.d.f. $P_n(x)$ tend towards the $P(x)$ that corresponds to $Q(u)$. Second, the additivity, i.e. if $Q(u)$ is of the form $\alpha Q'(u) + \beta Q''(u)$ with $\alpha + \beta = 1$, then $P(x)$ is $\alpha P'(x) + \beta P''(x)$ with the $P'(x)$, $P''(x)$ corresponding to $Q'(u)$ and $Q''(u)$ respectively. The following three groups of examples will illustrate the application of the foregoing results.

a) Let us first consider a function of two excess values ξ_1, ξ_2 only

$$(46) \qquad x = \frac{n}{2} f = \frac{n}{2} (A\xi_1^2 + 2B\xi_1\xi_2 + C\xi_2^2)$$

where the matrix Ψ is given by $\Psi_{11} = A$, $\Psi_{12} = \Psi_{21} = B$, $\Psi_{22} = C$. The product matrix $\bar{P}\Psi$ is

$$(47) \qquad \begin{matrix} A\bar{P}_{11} + B\bar{P}_{12} & B\bar{P}_{11} + C\bar{P}_{12} \\ A\bar{P}_{21} + B\bar{P}_{22} & B\bar{P}_{21} + C\bar{P}_{22} \end{matrix}$$

and the determinant of $I - \lambda\bar{P}\Psi$

$$(48) \quad D(\lambda) = 1 - \lambda[A\bar{P}_{11} + 2B\bar{P}_{12} + C\bar{P}_{22}] + \lambda^2(AC - B^2)(\bar{P}_{11}\bar{P}_{22} - \bar{P}_{12}^2).$$

If λ_1, λ_2 are the two real roots of $D(\lambda) = 0$, the asymptotic probability density of x will be

$$(49) \qquad \frac{dP(x)}{dx} = \frac{1}{2\pi} \int \frac{e^{-uix}\,du}{\sqrt{\left(1 - \dfrac{ui}{\lambda_1}\right)\left(1 - \dfrac{ui}{\lambda_2}\right)}}\,.$$

We are particularly interested in the case that $\bar{P}$ is "complete," i.e. a matrix with all horizontal and vertical sums vanishing. Then $\bar{P}_{11} = \bar{P}_{12} = \bar{P}_{22} = \overline{p_1 p_2}$, the last term in (48) cancels out and the only Eigenwert is $\lambda_1 = 1/(A - 2B + C)\overline{p_1 p_2}$. Here, instead of (49) we have

$$(50) \qquad \frac{dP(x)}{dx} = \frac{1}{2\pi} \int \frac{e^{-uix}\,du}{\sqrt{1 - \dfrac{ui}{\lambda_1}}} = \sqrt{\frac{\lambda_1}{\pi}}\,\frac{e^{-\lambda_1 x}}{\sqrt{x}}$$

This is, with respect to $\sqrt{|x|}$ a Gauss distribution with the variance $|A - 2B + C|\,\overline{p_1 p_2}/2$.

If, in addition to the assumption that $\bar{P}$ is "complete" (i.e. in the present case that $p_{\nu 1} + p_{\nu 2} = 1$ for all ν) the further assumption is made that the two intervals I_1 and I_2 cover the whole range of the original chance variables x_1, x_2, x_3, $\cdots$, one would have also $\xi_1 + \xi_2 = 0$ and from (46)

$$x = \frac{n}{2} (A - 2B + C)\xi_1^2.$$

In this case, $\sqrt{|x|}$ is a linear statistical function and the Central Limit Theorem leads to the same result as that expressed in (50). It is seen, however, from our derivation, that (50) holds under wider conditions: If $p_{\nu 1} + p_{\nu 2} = 1$ for all ν, there may exist another interval I_3 within the range of the chance variables x_1, x_2, x_3, $\cdots$ so that $\xi_1 + \xi_2$ is not necessarily zero.

The latter remark suggests the following general theorem: If f is a function of the k variables ξ_1, ξ_2, $\cdots$, ξ_k and g another such function but vanishing when $\xi_1 + \xi_2 + \cdots + \xi_k = 0$, then f and $f + g$ have the same asymptotic distribution provided that for each ν the sum $p_{\nu 1} + p_{\nu 2} + \cdots + p_{\nu k} = 1$. In the case of quadrics this result is equivalent to the following matrix theorem: If $\bar{P}$, Ψ, A are symmetric matrices, $\bar{P}$ with all horizontal and vertical sums equal to zero, Ψ arbitrary, and A of the form $a_{\iota\kappa} = a_\iota + a_\kappa$ then the two products

(51) $$\bar{P}\Psi \quad \text{and} \quad \bar{P}(\Psi + A)$$

have the same characteristic roots.—This can be proved by the usual methods of matrix calculus. The matrix $\bar{P}A$ has all characteristic roots equal to zero.[2]

b) In the definition of Karl Pearson's test function which is usually called χ^2, it is presumed that a sample is drawn from the combination of n *equal* distributions. In this case all $P^{(\nu)}$ are equal and coincide with $\bar{P}$ which then can simply be written P:

(52) $$P_{\iota\kappa} = p_\iota\delta_{\iota\kappa} - p_\iota p_\kappa.$$

The chance variable we now consider will be

(53) $$x = \frac{n}{2} f = \frac{n}{2} \sum_\iota \frac{\xi_\iota^2}{p_\iota} = \tfrac{1}{2}\chi^2.$$

Thus $\psi_{\iota\kappa} = \delta_{\iota\kappa}/p_\iota$ and the elements of $P\Psi$ are

(53') $$(P\Psi)_{\iota\kappa} = \sum_\mu P_{\iota\mu}\psi_{\mu\kappa} = \delta_{\iota\kappa} - p_\iota.$$

The matrix $I - \lambda P\Psi$ has the elements

$$\delta_{\iota\kappa}(1 - \lambda) + \lambda p_\iota.$$

If the kth column is subtracted from any one of the others, only two terms remain, one equal to $1 - \lambda$ and one equal $-(1 - \lambda)$ in the last row. Thus, the

---

[2] A proof of the matrix theorem has meanwhile been published by Alfred Brauer, *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 605–607.

determinant $D(\lambda)$ includes $(k - 1)$ times the factor $(1 - \lambda)$. On the other hand, $D(\lambda)$ is of degree $(k - 1)$ and has the absolute term 1. Therefore

$$(54) \qquad\qquad\qquad D(\lambda) = (1 - \lambda)^{k-1}.$$

This supplies the χ^2-distribution with $(k - 1)$ "degrees of freedom"

$$(55) \qquad Q(u) = (1 - ui)^{-\frac{k-1}{2}}, \quad \frac{dP(x)}{dx} = \frac{1}{\Gamma\left(\dfrac{k-1}{2}\right)} x^{\frac{k-3}{2}} e^{-x}, \qquad (x \geqq 0).$$

Again, our result is slightly more general than that reached in the usual theory. It includes the case that in addition to the k intervals with the probabilities $p_1, p_2, \cdots, p_k$ (whose sum is 1) there are other intervals with probability zero. On the other hand, if to χ^2 a term of the form $n\Sigma(a_\iota + a_\kappa)\xi_\iota\xi_\kappa$ is added, this would not change the asymptotic distribution.

One may ask for other quadratic functions of $\xi_1, \xi_2, \cdots, \xi_k$ whose asymptotic distribution is given by (55). In particular, one might be interested in a generalization of χ^2 for the case of *unequal original distributions*. The answer can easily be given by introducing the cofactors of order $(k - 1)$ and of order $(k - 2)$ of the determinant $|\bar{P}_{\iota\kappa}|$. It was mentioned in sec. 4 of Part I that all cofactors of order $(k - 1)$—in the case of "complete" $\bar{P}$—have the same value. It may be denoted by Δ. The cofactor corresponding to the lines ι, κ and the columns λ, μ will be denoted by $\Pi_{\iota\kappa;\lambda\mu}$ with $\Pi = 0$ if $\iota = \kappa$ or $\lambda = \mu$. Then, if l is any one of the integrers $1, 2, \cdots, k$

$$(56) \qquad\qquad\qquad \psi_{\iota\kappa} = \frac{1}{\Delta} \Pi_{\iota l;\kappa l}; \qquad \iota, \kappa \neq l$$

is one possible solution. In fact, the product $\bar{P}\Psi$ has in this case the elements
$(\bar{P}\Psi)_{\iota\kappa} = \delta_{\iota\kappa}, \quad \text{for} \quad \iota, \kappa \neq l$

$$(57) \qquad\qquad\qquad\qquad = -1, \text{ " } \iota = l, \kappa \neq l$$

$$\qquad\qquad\qquad\qquad\qquad = 0, \quad \text{ " } \kappa = l$$

The determinant of $I - \lambda\bar{P}\Psi$ is then seen to equal $(1 - \lambda)^{k-1}$.

The solution (56), however, is unsymmetrical in the sense that it does not include any terms with ξ_l. A completely symmetrical solution in which all ξ play the same role is given by

$$(58) \qquad\qquad\qquad \psi_{\iota\kappa} = \frac{1}{k\Delta} \sum_{l=1}^{k} \Pi_{\iota l;\kappa l}$$

According to (57) the matrix $\bar{P}\Psi$ now consists of terms $(k - 1)/k$ in the principal diagonal and $-1/k$ at all other places, that is

$$(58') \qquad\qquad\qquad (P\Psi)_{\iota\kappa} = \delta_{\iota\kappa} - \frac{1}{k}.$$

In the same way as in the case of (53′) it can be seen that the determinant of $I - \lambda P\Psi$ equals here $(1 - \lambda)^{k-1}$. *The asymptotic distribution of* $\sum \psi_{\iota\kappa}\xi_\iota\xi_\kappa$ *with the coefficients* (58) *is, therefore, the* χ^2-*distribution with* $(k - 1)$ *degrees of freedom.*

If the formula (58) is applied to the case of equal $P^{(\nu)}$ the corresponding quadric becomes

$$\sum_\iota \frac{1}{p_\iota} \xi_\iota^2 + \frac{1}{k} \sum_\iota \frac{1}{p_\iota} \left(\sum_\iota \xi_\kappa \right)^2,$$

that is, $\chi^2 +$ a term vanishing with $\xi_1 + \xi_2 + \cdots + \xi_k$. One can easily modify (58) so that it leads to χ^2 without any addition.

c) A third group of examples where the asymptotic density is expressed by simple functions is that where $D(\lambda)$ is an exact square, that is, all characteristic roots (except the one that is zero) have even multiplicities. Let us assume $k = 2m + 1$ and let $\lambda_1, \lambda_2, \cdots, \lambda_m$ be m double roots. Then

$$(59) \qquad Q(u) = \prod_{\mu=1}^m \left(1 - \frac{ui}{\lambda_\mu} \right)^{-1} = \sum_{\mu=1}^m \frac{\lambda_\mu A_\mu}{\lambda_\mu - ui}$$

with

$$(59) \qquad A_\mu = \prod_{\iota \neq \mu}^{1 \cdots m} \left(1 - \frac{\lambda_\mu}{\lambda_\iota} \right)$$

and therefore

$$(60) \qquad \frac{dP(x)}{dx} = \sum_{\mu=1}^m A_\mu \lambda_\mu e^{-\lambda_\mu x}, \qquad\qquad x \geq 0.$$

Assume, for instance, that all original distributions are uniform, that is

$$P_{\iota\kappa}^{(\nu)} = P_{\iota\kappa} = \frac{1}{k} \delta_{\iota\kappa} - \frac{1}{k^2}$$

and that the quadric f is given in the form (11) with the following $\alpha_{\iota\kappa}$:

$$(61) \qquad \begin{aligned} \alpha_{\iota\kappa} &= \sqrt{kc_1} & &\text{for } \iota = 1 \\ &= \sqrt{kc_\iota} & &\text{''} \quad \iota > 1, \kappa = 1, 2, \cdots, \iota - 1 \\ &= -(\iota - 1)\sqrt{kc_\iota} & &\text{''} \quad \iota > 1, \kappa = \iota \\ &= 0 & &\text{''} \quad \iota > 1, \kappa = \iota + 1, \iota + 2, \cdots, k. \end{aligned}$$

Then, the $\gamma_{\iota\kappa}$ as defined in (33) become

$$(62) \qquad \begin{aligned} \gamma_{\iota\kappa} &= c_\iota\iota(\iota - 1)\delta_{\iota\kappa} & &\text{for } \iota \text{ or } \kappa > 1 \\ &= 0 & &\text{''} \quad \iota = \kappa = 1 \end{aligned}$$

and $D(\lambda)$ according to (36) takes the form

$$(63) \qquad D(\lambda) = |\delta_{\iota\kappa} - \lambda\gamma_{\iota\kappa}| = \prod_{\iota=2}^k [1 - \lambda c_\iota \iota(\iota - 1)].$$

In other terms, for the quadric

$$f = kc_1(\xi_1 + \cdots + \xi_k)^2 + kc_2(\xi_1 - \xi_2)^2 + kc_3(\xi_1 + \xi_2 - 2\xi_3)^2 + \cdots$$
$$+ kc_k[\xi_1 + \xi_2 + \cdots + \xi_{k-1} - (k - 1)\xi_k]^2$$

the characteristic λ-values are $1/c_\iota \iota(\iota - 1)$.

Now, to obtain the case of m double roots with $k = 2m + 1$ we have simply to choose

$$c_2 = 3c_3 , \, 3c_4 = 5c_5 , \, 5c_6 = 7c_7 , \cdots .$$

The first term on the right-hand side can be entirely omitted in accordance to what was said in connection with (51). Besides, for the same reason, the expression can be simplified in various ways by assuming $\xi_1 + \xi_2 + \cdots + \xi_k = 0$.

As a numerical example, take $k = 5$, $c_2 = 3$, $c_3 = 1$, $c = 5$, $c_5 = 3$. Then

$$f = 20(\xi_1^2 + \xi_2^2 + \xi_3^2 + 20\,\xi_4^2 + 20\,\xi_5^2 - \xi_1\xi_2 - \xi_2\xi_3 - \xi_3\xi_1 + 10\,\xi_4\xi_5)$$

leads to the characteristic values $\lambda = 1/6$ and $1/60$ and the asymptotic density becomes

$$\frac{dP}{dx} = \frac{1}{54}\,(e^{-x/60} - e^{-x/6}).$$

In a similar way other groups of quadrics with asymptotic distributions of the type (60) can easily be constructed. One may, for instance, use eq. (41) and make vanish, in the matrix $S = \overline{P}\Psi$, all elements on one side of the diagonal so that the roots are immediately known.

5. Transition to the continuous case. In this concluding section, the transition to the case of a quadric of the form (1) with continuous $\psi\ (x, y)$ will be outlined. The formula best fit for this purpose is eq. (36). We therefore suppose the statistical function f given as

(64) $\quad f = \int\int \psi(x, y)\, dT_n(x)\, dT_n(y) \quad$ with $\quad \psi(x, y) = \int \alpha(r, x)\alpha(r, y)\, dr.$

In analogy to (33) we derive

(65)
$$\gamma(x, y) = \int\int \alpha(x, s)\alpha(y, t)\, d\overline{U}_n(s, t)$$
$$= \int \alpha(x, s)\alpha(y, s)\, d\overline{V}_n(s) - \int\int \alpha(x, s)\alpha(y, t)\, d\overline{W}_n(s, t).$$

Since $d\overline{W}$ is symmetric, this function $\gamma(x, y)$ is symmetric with respect to x and y. If $D(\lambda)$ denotes the *Fredholm determinant* of the "kernel" $\gamma(x, y)$, we con-

clude from (36) that the characteristic function of the asymptotic distribution of f will be given by

$$(66) \qquad Q_n(u) \sim \frac{1}{D(ui)}$$

if certain convergence conditions are satisfied.

In order to establish (66) the main point is to find a sequence of functions $\psi_1(x, y), \psi_2(x, y), \cdots$ each of the type considered in the foregoing Sections and such that 1) the distribution of the quadric f_k with the coefficients ψ_L tends towards the distribution of f with increasing k and independently of n; and 2) that the determinants D_k corresponding to ψ_k converge towards D as k increases indefinitely. Using our Lemma A we can replace the first condition by asking that the expectation of $| f - f_k |$ should go to zero with $k \to \infty$ independently of n.

The following assumptions shall be made concerning f and the $V_r(x)$: The function $\alpha(r, x)$ in (64) is continuous and bounded in every finite region; there exist two positive continuous functions $\alpha(r), \beta(x)$ such that

$$(67) \qquad | \alpha(r, x) | \leq \alpha(r)\beta(x)$$

and that the integrals

$$(68) \qquad \int \alpha^2(r) \, dr = M, \qquad \int \beta(x) \, dV_r(x), \qquad \int \beta^2(x) \, dV_r(x)$$

exist, the latter two being bounded and converging uniformly with respect to ν. We are going to devise a step function $\psi_k(x, y)$ so that for the corresponding f_k and any positive ϵ_1

$$(69) \qquad E\{ | f - f_k | \} \leq \epsilon_1 .$$

Let N be an upper bound of the integrals

$$(70) \qquad \int \beta(x) \, dV_r(x) \leq N, \qquad \int \beta(x) \, d\bar{V}_n(x) \leq N$$

and $\epsilon = \epsilon_1/(5 + 8 N)$. Choose a value L such that

$$(71) \qquad \int_{|x| > L} \beta(x) \, dV_r(x) \leq \frac{\epsilon}{M}, \qquad \int_{|x| > L} \beta^2(x) \, dV_r(x) \leq \frac{\epsilon}{M}$$

and, calling B the maximum of $\beta(x)$ in $| x | \leq L$, another quantity R such that

$$(72) \qquad \int_{|r| > R} \alpha^2(r) \, dr \leq \frac{\epsilon}{2B^2} .$$

We subdivide, in the x-y-r-space, the domain $| x | \leq L, | y | \leq L, | r | \leq R$ in k^3 equal cells where k is determined by the condition that the absolute value of the variation of $\alpha(r, x)\alpha(r, y)$ within each cell does not exceed $\epsilon/4R$. Outside this domain we set $\psi_k(r, x) = 0$ while inside the domain $\alpha_k(r, x)\alpha_k(r, y)$ shall

equal the value that $\alpha(r, x)\alpha(r, y)$ assumes in the center of the respective cell. Then $\psi_k(x, y)$ will be defined by

(73) $$\psi_k(x, y) = \int \alpha_k(r, x)\alpha_k(r, y)\, dr.$$

From the definition of k and from (67) and (72) it follows that

(74)
$$| \psi(x, y) - \psi_k(x, y)| \leqq \int_{|r| \leqq R} | \alpha(r, x)\alpha(r, y) - \alpha_k(r, x)\alpha_k(r, y)|\, dr$$
$$+ \int_{|r| > R} | \alpha(r, x)\alpha(r, y)|\, dr$$
$$\leqq 2R\,\frac{\epsilon}{4R} + \beta(x)\beta(y) \int_{|r| > R} \alpha^2(r)\, dr \leqq \frac{\epsilon}{2} + B^2\,\frac{\epsilon}{2B^2} = \epsilon$$

as long as $| x | \leqq L, | y | \leqq L$. If this square is called (L) and the complementary region $(\bar{L})$ we have

(75)
$$f - f_k = \iint_{(L)} [\psi(x, y) - \psi_k(x, y)]\, dT_n(x)\, dT_n(y)$$
$$+ \iint_{(\bar{L})} \psi(x, y)\, dT_n(x)\, dT_n(y)$$

and since the integral of $| dT_n(x)\, dT_n(y) |$ is not larger than 4, while, according to (64) and (67)

(76) $$| \psi(x, y) | \leqq \beta(x)\beta(y) \int \alpha^2(r)\, dr = M\beta(x)\beta(y)$$

we conclude from (74) and (75)

(77) $$|f - f_k| \leqq 4\epsilon + M \iint_{(\bar{L})} \beta(x)\beta(y)\, | dT_n(x)\, dT_n(y) |.$$

This gives

(78) $$E\{|f - f_k|\} \leqq 4\epsilon + M \iint_{(\bar{L})} \beta(x)\beta(y)E\{| dT_n(x)\, dT_n(y) |\}.$$

Now, from $| dT_n | = | dS_n - d\bar{V}_n | \leqq dT_n + 2d\bar{V}_n$ and from the formulas derived in Part II,

$$E\{ dT_n(x)\} = 0, \qquad E\{ dT_n(x)\, dT_n(y)\} = \frac{1}{n}\, d\bar{U}_n(x, y)$$

it follows

(79) $$E\{ | dT_n(x)\, dT_n(y) | \} \leqq \frac{1}{n}\, d\bar{U}_n(x, y) + 4\, d\bar{V}_n(x)\, d\bar{V}_n(y)$$

with ·

(79′) $d\bar{U}_n(x, y) = \delta(x, y) \, d\bar{V}_n(x) - d\bar{W}_n(x, y) \leqq \delta(x, y) \, d\bar{V}_n(x).$

If this is introduced in (78) and (71) taken into account, we find

$$E\{ |f - f_k| \} \leqq 4\epsilon + M \frac{1}{n} \int_{|x| > L} \beta^2(x) \, d\bar{V}_n(x)$$

(80) $+ 4M \iint_{(L)} \beta(x)\beta(y) \, d\bar{V}_n(x) \, d\bar{V}_n(y)$

$$\leqq 4\epsilon + \frac{1}{n} \epsilon + 4 \times 2N\epsilon \leqq (5 + 8N)\epsilon = \epsilon_1$$

as required in (69).

On the other hand, it can be seen that the kernel $\gamma(x, y)$ as defined in (65) is the limit of the sequence $\gamma_k(x, y)$

(81)
$$\gamma_k(x, y) = \iint_{(L)} \alpha_k(x, s)\alpha_k(y, t) \, d\bar{U}_n(s, t) \qquad \text{for } x, y \text{ in } (R)$$

$$= 0 \qquad \text{for } x, y \text{ in } (\bar{R})$$

where (R) means the region $|x| \leqq R, |y| \leqq R$ and $(\bar{R})$ the complementary region. In fact, from the definition of k and eqs. (67) and (71) one has for x, y in (R):

$$| \gamma(x, y) - \gamma_k(x, y) | \leqq \frac{\epsilon}{4R} \iint_{(L)} | d\bar{U}_n(s, t) |$$

$$+ \iint_{(L)} | \alpha(x, s)\alpha(y, t) \, d\bar{U}_n(s, t) |$$

(82) $\leqq \dfrac{\epsilon}{2R} + \alpha(x)\alpha(y) \left[\displaystyle\int_{|s| > L} \beta^2(s) \, d\bar{V}_n(s) \right.$

$$\left. + \frac{1}{n} \sum_{\nu=1}^{n} \iint_{(L)} \beta(s)\beta(t) \, dV_\nu(s) \, dV_\nu(t) \right.$$

$$\leqq \frac{\epsilon}{2R} + \alpha(x)\alpha(y) \frac{\epsilon}{M} (1 + 2N).$$

Since $\alpha(x)$ is bounded, the right-hand side goes to zero with ϵ. Finally, for x, y in $(\bar{R})$ we have

$$| \gamma(x, y) - \gamma_k(x, y) | \leqq \iint | \alpha(x, s)\alpha(y, t) \, d\bar{U}_n(s, t) |$$

(83) $\leqq \alpha(x)\alpha(y) \left[\displaystyle\int \beta^2(s) \, d\bar{V}_n(s) \right.$

$$\left. + \frac{1}{n} \sum_{\nu=1}^{n} \iint \beta(s)\beta(t) \, dV_\nu(s) \, dV_\nu(t). \right]$$

Here, the two terms in the brackets are bounded, but $\alpha(x)\alpha(y)$ goes to zero as R increases. The conclusion is that $\gamma_k(x, y)$ tends uniformly towards $\gamma(x, y)$ with $k \to \infty$.

Thus, eq. (66) is established provided that the function $\gamma(x, y)$ defined in (65) has a Fredholm determinant $D(\lambda)$ that is the limit of the corresponding algebraic determinants and provided that the c.f. $\sqrt{1/D(ui)}$ leads to a c.d.f. with bounded derivative.

As an *example* let us consider the case

$$
\begin{aligned}
(84) \qquad \alpha(r, x) &= \sqrt{g'(r)} \text{ for } r \geqq x \\
&= 0 \qquad `` \ r < x.
\end{aligned}
$$

This function is not continuous as it was assumed in establishing (66). However, the existence of a single discontinuity line, $x = r$, does not invalidate the argument. We assume $g'(r) = 0$ and equal to dg/dr. Then, in the case of (84):

$$
\begin{aligned}
(85) \qquad \psi(x, y) &= \int \alpha(r, x)\alpha(r, y)dr = -g(y) \quad \text{for } x \leqq y \\
&= -g(x) \quad `` \ x \geqq y.
\end{aligned}
$$

Since, however, adding to ψ a function of x or of y alone does not change the value of f, we can also use

$$
\begin{aligned}
(85') \qquad \psi(x, y) &= g(x) \quad \text{for } x \leqq y \\
&= g(y) \quad `` \ x \geqq y.
\end{aligned}
$$

The statistical function f that corresponds to (84) can be computed either from (85) or (85')—or directly from (84) if we use the formula that follows from (64)

$$
(86) \qquad f = \int \left[\int \alpha(r, x) \, dT_n(x) \right]^2 dr.
$$

The integral in the brackets is, in our case, seen to equal $\sqrt{g'(r)} \ T_n(r)$, thus

$$
(86') \qquad f = \int g'(r)[S_n(r) - \bar{V}_n(r)]^2 \, dr.
$$

This is exactly the test function ω^2 mentioned in the Introduction, eq. (3).

To find the distribution of f we have to compute $\gamma(x, y)$. Its definition (65) can be written in the form

$$
(87) \quad \gamma(x, y) = \frac{1}{n} \sum_{r=1}^{n} \left[\int \alpha(x, s)\alpha(y, s) \, dV_r(s) - \int \alpha(x, s) \, dV_r(s) \int \alpha(y, s) \, dV_r(s) \right].
$$

This supplies in the case of (84)

$$
\begin{aligned}
(88) \qquad \gamma(x, y) &= \sqrt{g'(x)g'(y)}[\bar{V}_n(x) - \overline{V_n(x)V_n(y)}] \text{ for } x \leqq y \\
&= \sqrt{g'(x)g'(y)}[\bar{V}_n(y) - \overline{V_n(x)V_n(y)}] \quad `` \ x \geqq y.
\end{aligned}
$$

Here, the second term in the brackets is the arithmetical mean of the products $V_\nu(x)V_\nu(y)$.

If the distributions $V_\nu(x)$ are all equal (independent of ν) we have simply to write $V(x)$ instead of $\bar{V}_n(x)$ and $V(x)V(y)$ instead of $\overline{V_n(x)V_n(y)}$. If, in addition, the distribution in the original collectives are uniform in the basic interval 0 to 1, one has

$$(89) \qquad \begin{aligned} \gamma(x, y) &= \sqrt{g'(x)g'(y)}\, x\,(1 - y) \text{ for } 0 \leq x \leq y \leq 1 \\ &= \sqrt{g'(x)g'(y)}\, y(1 - x) \quad `` \ 0 \leq y \leq x \leq 1. \end{aligned}$$

This is the case dealt with in Smirnoff's papers [7, 8]. If, finally, $g'(x)$ is supposed to be equal to 1 in the interval 0, 1, we arrive at a kernel $\gamma(x, y)$ whose Fredholm determinant is well known:

$$(90) \qquad \begin{aligned} \gamma(x, y) &= x(1 - y) \quad \text{ for } \quad x \leq y \\ &= y(1 - x) \quad `` \quad x \geq y. \end{aligned} \qquad D(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}}.$$

This supplies immediately the c.f. and (in form of a definite integral) the c.d.f. of the asymptotic distribution of ω^2 for $g' = 1$.

The same result can be reached without the use of $\alpha(r, x)$ if we apply one of the transformations discussed in the foregoing Section. Take, for instance, instead of $\gamma(x, y)$ the unsymmetric kernel $\sigma(x, y)$ corresponding to the matrix $S = P\Psi$ defined in (41). If all original distributions are equal, the element of S can be written as

$$(91) \qquad s_{\iota\kappa} = \sum_\mu P_{\iota\mu}\psi_{\mu\kappa} = p_\iota(\psi_{\iota\kappa} - \sum_\mu \psi_{\mu\kappa}p_\mu).$$

Calling $\iota(x)$ the density $dV(x)/dx$ in the continuous case, the corresponding kernel becomes

$$(92) \qquad \sigma(x, y) = v(x)\left[\psi(x, y) - \int \psi(s, y)v(s)\, ds\right].$$

With the ψ-values from (85'), $g' = 1$, $v = 1$, this gives

$$(92') \qquad \begin{aligned} \sigma(x, y) &= x - y + \frac{y}{2} \text{ for } \quad x \leq y \\ &= \frac{y^2}{2} \qquad `` \qquad x \geq y. \end{aligned}$$

It can easily be seen that the "Eigenfunctions" of this $\sigma(x, y)$ are $\sin(\sqrt{\lambda_m}\, x)$ with $\lambda_m = m^2\pi^2$, and, therefore, the Fredholm determinant is that indicated in (90).

It might be added that the expectation and the asymptotic variance of ω^2 can be computed, independently of the distribution, from the formulas developed in Part I. The results are

$$(93) \qquad nE\{\omega^2\} = \int g'(x)\overline{V_n(x)[1 - V_n(x)]}\, dx$$

and, in the case of all $V_\nu(x)$ equal

$$(94) \qquad n^2 \text{Var}\{\omega^2\} \sim 4 \iint\limits_{x \leq y} g'(x)g'(y) V^2(x)[1 - V(y)]^2 \, dx \, dy.$$

These formulas have already been given in [4].

Another, more general, remark is this. If all $V_\nu(x)$ are equal, one can reduce the problem, by a transformation of the original chance variable x into $x' = V(x)$, to the case of a uniform distribution over the interval 0 to 1. If the $V_\nu(x)$ are not equal, it might still be possible to find a transformation $x' = x'(x)$ such that all original distributions extend over a finite region on the x'-axis only. In this case the restrictions concerning the behavior of the distributions at infinity drop out.

REFERENCES

[1] HARALD CRAMÉR, "On the composition of elementary errors," *Skand. Aktuarietids-drift*, Vol. 11 (1928), pp. 13–74, 141–180.

[2] R. V. MISES, "Les lois de probabilité pour les fonctions statistiques," *Ann. de l'Inst. Henri Poincaré*, Vol. 6 (1936), pp. 185–212.

[3] ————, "Sur les fonctions statistiques," *Soc. math. de France*, Conférence de la Réunion internat. des Mathématiciens, Paris, 1937.

[4] ————, *Wahrscheinlichkeitsrechnung und ihre Anwendung*, Leipzig and Wien, 1931.

[5] G. PÓLYA, "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung," *Math. Zeitschr.*, Vol. 8 (1920), pp. 171–181.

[6] T. A. SHOHAT AND T. D. TAMARKIN, *The Problem of Moments*, Math. Surveys No. 1, New York, 1943.

[7] N. V. SMIRNOFF, "On the distribution of the ω^2-criterion of Mises," (In Russian), *Recueil Math.*, nouvelle série, Vol. 2 (1937), pp. 973–993.

[8] ————, "Sur la distribution de ω^2 (criterium de M. von Mises)," *Comptes Rendus Paris*, Vol. 202 (1936), p. 449.

[9] VITO VOLTERRA. *Leçons sur les Fonctions de Ligne*, Paris, 1913.

[10] VITO VOLTERRA AND JÓSEPH PÉRÈS, *Théorie générale des Fonctionelles*, Paris, 1936.

APPROXIMATE SOLUTIONS FOR MEANS AND VARIANCES IN A CERTAIN CLASS OF BOX PROBLEMS

By Philip J. McCarthy

Social Science Research Council

1. Summary. Consider n boxes, each box having an associated probability, p_i, $(\sum_i p_i = 1)$, and an associated integer, k_i. If balls are thrown one by one into these boxes, the probability being p_i that any one ball falls into the ith box, then the number of balls which must be thrown in order to obtain, *for the first time*, at least k_{i_1} balls in the i_1th box, at least k_{i_2} balls in the i_2th box, $\cdots$, and at least k_{i_s} balls in the i_sth box, is a random variable, $N_s[k_1(p_1), k_2(p_2), \cdots, k_n(p_n)]$. Here i_1, i_2, $\cdots$, i_s represent the numbers of that set of s boxes, $(1 \leq s \leq n)$, which *first* satisfies the stated condition.

The distribution of $N_s[k_1(p_1), k_2(p_2), \cdots, k_n(p_n)]$ can be written down for any set of values assigned to n, s, the p_i's and the k_i's. However, for n greater than 2 the distribution assumes such an extremely complicated multinomial form that except for certain special cases even the mean of the distribution cannot be numerically evaluated without a prohibitive amount of labor.

This paper presents the exact moments of $N_1[k_1(p_1), k_2(p_2)]$ and $N_2[k_1(p_1), k_2(p_2)]$ in forms that readily lend themselves to computation and shows how these moments can be used to obtain approximate values for the mean and variance for certain situations where n is greater than two. These approximation formulae are given for

1. The mean and variance, for any n and any set of k_i's and p_i's when $s = 1$ or n.

2. The mean, for any n and $2 \leq s \leq n - 1$, when $p_i = 1/n$, $k_i = k$, $(i = 1, 2, \cdots, n)$.

Some indications are given concerning the error of the approximations, and the circumstances which lead to a minimum (and maximum) error. Curves have been prepared to show the mean for the two box case, the primary function of these curves being to assist in the application of the approximation formulae. Some problems where the results of this paper might be applicable are suggested in the Introduction.

2. Introduction. A box problem is defined when one is given a fixed number of boxes, a collection of balls (either finite or infinite), a set of rules governing the throwing of the balls into the boxes and a statement of the conditions which will bring the throwing to an end. The terminating conditions usually state either that a fixed number of balls will be thrown or that balls will be thrown until a particular distribution of balls in the boxes has been obtained. In the first of these, interest is centered on the possible distributions which can be ob-

tained, while in the latter the number of balls necessary to obtain a specified distribution is of primary interest.

This paper will be concerned with certain problems falling in the latter category. In the simplest case one is given two boxes with associated probabilities p_1 and p_2 and associated integers k_1 and k_2. Balls are thrown one by one into the two boxes, the probability being p_1 that any one ball goes in the first box and p_2 that it goes in the second box. This process is stopped when either k_1 balls fall in box 1 or k_2 balls in box 2, *whichever occurs first*. One is interested in the distribution of the number of balls necessary to terminate the throwing. This problem was stated in essentially this form by Laplace [4], but he contented himself with merely writing down the probability generating function.

Here the special case of two boxes will be treated in detail and the results will then be generalized to the n-box case. In all of these instances it is possible to write down exact expressions for the mean and variance of the number of balls required to achieve the stated distribution. However, in almost every case the resulting expressions are too complicated to be of any use when a numerical answer is desired. The principal portion of this paper will be devoted to obtaining approximate formulae from which numerical answers can be obtained for these problems. Some evaluation of the degree of approximation will be given in section 5, while curves to facilitate the computation will be given in section 6.

The statement of these problems in terms of boxes and balls may lead one to the belief that they have no other interpretation. Actually this is not the case, and a few illustrations of this point will now be given. For example, consider the curtailed single sampling plan used in acceptance sampling. A buyer receives a lot of articles. This lot will contain a certain proportion of defective items. The buyer wishes to determine on the basis of sampling whether to accept or reject the lot. His knowledge of his own situation will allow him to specify the largest proportion of defectives which he is ordinarily willing to accept and the risk he is willing to take of accepting a lot with a proportion defective larger than this critical proportion. On the basis of these two values it is possible to set up a sampling plan in which the buyer will take a sample of size n out of the lot, inspect it, and reject it if there are k_1 or more defectives in the sample. Of course once he has obtained k_1 defectives there is no need to inspect the remainder of the sample. The lot will then be automatically rejected. Similarly, once he has obtained $n - k_1$ non-defectives, he can accept the lot without inspecting the remainder of the items. The average number of items which he must inspect in order to reach a decision is given by the solution to the two box problem stated above. Box 1 will receive the defective items, the associated integer being k_1 and the associated probability being p_1, the true proportion of defectives in the lot. Box 2 will receive the non-defective items, the associated integer being $n - k_1$ and the associated probability being p_2, the true proportion of non-defectives in the lot.

Laplace [4] considered problems of this type as applied to games of chance.

Thus suppose there are two players A and B who participate in successive trials of a given event, the probability being p_1 that A wins on any one trial and p_2 that B wins. Then one can associate the integer k_1 with A and k_2 with B by saying that A wins the match if he wins k_1 trials before B wins k_2 trials and conversely. The analysis is exactly the same as for the two box problem. It is apparent that this same situation can be extended to any number of players.

Another possible interpretation is as a particular kind of random walk problem. Let a particle start at the origin of a system of rectangular coordinates and suffer successive positive unit displacements, the probability being p_1 that it moves one unit in the x-direction and p_2 that it moves one unit in the y-direction. Furthermore assume that it is absorbed if it ever reaches the line $x = k_1$ or the line $y = k_2$. Then the analysis of the above two box problem gives the mean number of displacements before it is absorbed. In the same manner, such a random walk problem can be stated for n dimensions. For n equal to three, there will be three planes and the particle will be absorbed when it reaches any one of the three.

Certain problems in public opinion polling may fit into this category of box problems, particularly if the above problem is rephrased so that one requires the mean number of trials to obtain at least k_1 balls in the first box and at least k_2 balls in the second box, *for the first time*. For example, suppose that one desires to sample from a population composed of two types of individuals, A and B. Let the population proportions of A and B be known and be denoted by p_1 and p_2. Then if one wishes to obtain at least k_1 individuals of type A and at least k_2 individuals of type B, the average number of persons who must be chosen in order to fulfill this condition is given by the analysis of the corresponding box problem. This is rather artificial when there are only two categories and $p_1 + p_2 = 1$. However, these restrictions will be removed in the course of the paper, and the problem will be considered for any number of types of individuals.

As a final example, consider one of the many bombing problems which arose during the course of war research. Suppose that a factory which is to be demolished has n vital units, the destruction of any one of which will destroy the usefulness of the factory. Let the probability be p_1 of hitting the first unit with a single bomb, p_2 the probability of hitting the second with a single bomb, etc., and assume that k_1 bomb hits will finish off the first unit, k_2, the second, etc. Then the mean number of bombs required will be given by the analysis for the corresponding box problem.

Corresponding interpretations are possible for the other problems which are to be considered in this paper. Some of these will be indicated as the analysis proceeds and it is to be hoped that others will occur to the reader.

As previously noted, this paper will be concerned with the distribution of balls necessary to terminate the throwing, assuming the p's are known. Another possible interpretation is to assume the p's unknown and to estimate them with the results of the ball throwing. Certain aspects of this problem for two boxes

have been considered by J. B. S. Haldane [3] and Girshick, Mosteller and Savage [2].

3. Solution for the two box case.

3.1. *Distribution and moments of the number of trials necessary to obtain either* k_1 *balls in the first box or* k_2 *balls in the second box.* This problem may be stated as follows: Suppose one is given two boxes with associated probabilities p_1 and p_2, and associated integers k_1 and k_2. For the present it will be assumed that $p_1 + p_2 = 1$, although this restriction will be removed later. Now let balls be thrown one by one into these two boxes, the probability being p_1 that a particular ball will fall in the first box and p_2 that it will fall in the second box. This process is stopped on the first ball which leaves either k_1 balls in the first box or k_2 balls in the second box. The number of balls, x, which is required to accomplish this is a random variable and we desire the moments of x. The probability that k_1 balls are obtained in the first box on the xth throw, $k_1 \leq x \leq k_1 + k_2 - 1$, before k_2 balls are obtained in the second box, is immediately seen to be

$$(3.1) \qquad \left[\binom{x-1}{k_1-1} p_1^{k_1-1} p_2^{x-k_1} \right] \cdot p_1 = \binom{x-1}{k_1-1} p_1^{k_1} p_2^{x-k_1} .$$

Similar reasoning gives the probability that k_2 balls are obtained in the second box for the first time on the xth throw, $k_2 \leq x \leq k_1 + k_2 - 1$, as

$$(3.2) \qquad \binom{x-1}{k_2-1} p_1^{x-k_2} p_2^{k_2} .$$

From (3.1) and (3.2), the hth moment of x, $E(x^h)$, is

$$(3.3) \qquad \sum_{x=k_1}^{k_1+k_2-1} x^h \binom{x-1}{k_1-1} p_1^{k_1} p_2^{x-k_1} + \sum_{x=k_2}^{k_1+k_2-1} x^h \binom{x-1}{k_2-1} p_1^{x-k_2} p_2^{k_2} .$$

However, it is inconvenient to consider (3.3) directly. A much simpler procedure is to determine the increasing factorial moments of x and then transform these into the ordinary moments. Thus the hth increasing factorial moment of x, $F_{h,1}[k_1(p_1), k_2(p_2)]$, is defined as $E[x(x+1) \cdots (x+h-1)]$. Then $F_{h,1}[\ \ ]$ is equal to

$$(3.4) \qquad \sum_{x=k_1}^{k_1+k_2-1} \frac{(x+h-1)!}{(x-1)!} \binom{x-1}{k_1-1} p_1^{k_1} p_2^{x-k_1}$$
$$+ \sum_{x=k_2}^{k_1+k_2-1} \frac{(x+h-1)!}{(x-1)!} \binom{x-1}{k_2-1} p_1^{x-k_2} p_2^{k_2} .$$

(3.4) can be transformed by means of the relationship

$$(3.5) \qquad \sum_{j=0}^{a} \binom{k+j}{j} p_1^j = (1-p_1)^{-(k+1)} I_{1-p_1}(k+1, a+1),$$

where $I_x(p, q)$ is the Incomplete Beta-Function as tabulated by Karl Pearson [6], and the result is obtained that

$$
\begin{aligned}
(3.6) \quad F_{h,1}[k_1(p_1), k_2(p_2)] &= \frac{k_1(k_1 + 1) \cdots (k_1 + h - 1)}{p_1^h} I_{p_1}(k_1 + h, k_2) \\
&\quad + \frac{k_2(k_2 + 1) \cdots (k_2 + h - 1)}{p_2^h} I_{p_2}(k_2 + h, k_1).
\end{aligned}
$$

The ordinary hth moment of x may be written in terms of $F_{1,1}[\ \ ], F_{2,1}[\ \ ], \cdots,$ $F_{h,1}[\ \ ]$ as

$$
(3.7) \qquad E(x^h) = \sum_{i=1}^{h} F_{i,1}[\ ] \frac{\Delta^i 0^h}{i!} (-1)^{h+i},
$$

where $\Delta^i 0^h$ represents a difference of zero. Tabular values of $\Delta^i 0^h / i!$ are given by Fisher and Yates [1].

In particular, the mean and variance of x, which will receive the special designations $E_1[k_1(p_1), k_2(p_2)]$ and $\sigma_1^2[k_1(p_1), k_2(p_2)]$ respectively, are

$$
(3.8) \qquad \frac{k_1}{p_1} I_{p_1}(k_1 + 1, k_2) + \frac{k_2}{p_2} I_{p_2}(k_2 + 1, k_1)
$$

and

$$
\begin{aligned}
(3.9) \quad &\frac{k_1(k_1 + 1)}{p_1^2} I_{p_1}(k_1 + 2, k_2) + \frac{k_2(k_2 + 1)}{p_2^2} I_{p_2}(k_2 + 2, k_1) \\
&\quad - E_1[k_1(p_1), k_2(p_2)] - \{E_1[k_1(p_1), k_2(p_2)]\}^2.
\end{aligned}
$$

In the event the p's are equal and sum to one, $E_1[k_1(p_1), k_2(p_2)]$ will be abbreviated to $E_1[k_1, k_2]$, and finally, if both the p's and k's are equal, it will be written as $E_1[k^2]$. In this two box situation, the only other possibility is $E_2[k_1(p_1), k_2(p_2)]$, which will denote the expected number of balls required to obtain at least k_1 in the first box and at least k_2 in the second box, for the first time. This problem will be considered in section 3.2.

In order to facilitate the computation of mean values, both for the two box problem itself and for its application to problems involving a larger number of boxes, (3.8) has been graphed for various values of k_1, k_2, p_1 and p_2. A discussion of this procedure and the results obtained will be found in section 6.

There is one further result which will later prove useful. Consider the situation when there is only one box with p_1 and k_1, $p_1 < 1$. This is the same as having two boxes where the k_2 corresponding to the second box is infinite. In other words, one can terminate the throwing of balls only because of what happens to the first box, never because of anything that happens to the second box. In this case one obtains

$$
(3.10) \qquad E_1[k_1(p_1), \infty \ (p_2)] = \sum_{x=k_1}^{\infty} x \binom{x-1}{k_1-1} p_1^{k_1} p_2^{x-k_1} = \frac{k_1}{p_1}.
$$

Similarly,

$$(3.11) \qquad\qquad \sigma_1^2[k_1(p_1),\ \infty\ (p_2)] = \frac{k_1 p_2}{p_1^2}.$$

3.2. *Distribution and moments of the number of throws necessary to obtain at least k_1 balls in the first box and at least k_2 balls in the second box.* This problem may be stated as follows: Suppose there are two boxes with associated probabilities p_1 and p_2, and associated integers k_1 and k_2. As in 3.1, $p_1 + p_2 = 1$. Let balls be thrown into the boxes one by one, the probability being p_1 that a particular ball will fall in the first box and p_2 that it will fall in the second box. This process is stopped on the first ball which leaves at least k_1 in the first box and exactly k_2 in the second or at least k_2 in the second and exactly k_1 in the first. Again x is the number of balls required to accomplish this. As explained in 3.1, the mean value in this case will be written as $E_2[k_1(p_1), k_2(p_2)]$. The analysis follows through as in 3.1 and the mean number of trials is equal to

$$(3.12) \qquad \sum_{x=k_1+k_2}^{\infty} x \binom{x-1}{k_1-1} p_1^{k_1} p_2^{x-k_1} + \sum_{x=k_1+k_2}^{\infty} x \binom{x-1}{k_2-1} p_1^{x-k_2} p_2^{k_2}.$$

Making use of (3.5), this can be written as

$$(3.13) \qquad \frac{k_1}{p_1} [1 - I_{p_1}(k_1 + 1, k_2)] + \frac{k_2}{p_2} [1 - I_{p_2}(k_2 + 1, k_1)].$$

Referring to (3.8) it is evident that

$$(3.14) \qquad E_1[k_1(p_1), k_2(p_2)] + E_2[k_1(p_1), k_2(p_2)] = \frac{k_1}{p_1} + \frac{k_2}{p_2}.$$

The hth increasing factorial moment in this problem, denoted by $F_{h,2}[k_1(p_1), k_2(p_2)]$, is

$$(3.15) \qquad \frac{k_1(k_1+1)\cdots(k_1+h-1)}{p_1^h} [1 - I_{p_1}(k_1 + h, k_2)]$$
$$+ \frac{k_2(k_2+1)\cdots(k_2+h-1)}{p_2^h} [1 - I_{p_2}(k_2 + h, k_1)].$$

Comparison of (3.15) with (3.6) gives the relationship

$$(3.16) \qquad F_{h,1}[\ ] + F_{h,2}[\ ] = \frac{k_1(k_1+1)\cdots(k_1+h-1)}{p_1^h}$$
$$+ \frac{k_2(k_2+1)\cdots(k_2+h-1)}{p_2^h}.$$

The ordinary moments of x can be computed from (3.15) by the use of (3.7). That is, formula (3.7) holds in this case if $F_{h,1}[\quad]$ is replaced by $F_{h,2}[\quad]$.

It can be easily shown by the use of the recursion relationship for the Incomplete Beta-Function,

$$I_x(p, q) = xI_x(p - 1, q) + (1 - x)I_x(p, q - 1),$$

that $F_{h,1}[\ ]$ and $F_{h,2}[\ ]$ satisfy the partial difference equation

$$
\begin{aligned}
F_{h,i}[k_1(p_1), k_2(p_2)] = \ & hF_{h-1,i}[k_1(p_1), k_2(p_2)] \\
(3.17) \qquad\qquad & + p_1 F_{h,i}[(k_1 - 1)(p_1), k_2(p_2)] \\
& + p_2 F_{h,i}[k_1(p_1), (k_2 - 1)(p_2)],
\end{aligned}
$$

where $i = 1$ or 2. This equation can be used as an alternative way of obtaining many results, examples of which are (3.10) and (3.11). Certain of these applications have been discussed by McCarthy [5].

4. Solution for the n box case.

4.1. *Preliminary discussion.* The problems of this section, although direct generalizations of the two box cases, can perhaps be most easily stated and illustrated as applied to the behavior of a random particle. Suppose that we have a random particle which starts at the origin of n-dimensional rectangular coordinates and moves in unit steps along the positive coordinate axes. At any given point the probability will be taken as p_i that it moves in the x_i-direction. $\sum_{i=1}^{n} p_i$ is assumed to be one unless otherwise specified. Now consider the n hyperplanes, $x_i = k_i$, and assume that the particle will be absorbed if it passes through a *specified number*, say s, of these hyperplanes. Notice that we are interested only in the number of planes which it passes through, and not in the particular ones. For each s, $(s = 1, 2, \cdots, n)$, the number of moves which the particle makes before it is absorbed is a random variable, and in this section we will be concerned with the distribution of this random variable. The corresponding interpretations for boxes and balls is immediately obvious.

These problems are seen to be generalizations of the two box cases considered in section 3. Although it is always relatively easy to write down formal expressions for the quantities to be considered, the step from two boxes to three or more boxes produces expressions which are extremely difficult, or even impossible, to evaluate. In this section we shall develop approximate solutions which make use only of simple computations based on the solution for the two box case.

As an introduction to the contents of this section, we shall discuss briefly a box problem which is a special case of the general problem. Assume that there are n boxes with a probability of $1/n$ that any one ball will be thrown into a particular one of the n boxes. Then one can ask for the mean and variance of

the number of trials required to obtain s occupied boxes (i.e. $k_1 = k_2 = \cdots = k_n = 1$). Making use of (3.10) and (3.11), we obtain

$$E_1[1^n] = 1$$

$$E_2[1^n] = 1 + E_1\left[1\left(\frac{n-1}{n}\right),\ \infty\left(\frac{1}{n}\right)\right] = 1 + \frac{n}{n-1}$$

$$E_3[1^n] = 1 + \frac{n}{n-1} + E_1\left[1\left(\frac{n-2}{n}\right);\ \infty\left(\frac{2}{n}\right)\right]$$

(4.1)
$$= 1 + \frac{n}{n-1} + \frac{n}{n-2}$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$E_s[1^n] = 1 + \frac{n}{n-1} + \cdots + \frac{n}{n-s+1} = n\sum_{i=0}^{s-1}\frac{1}{n-i},$$

and

$$\sigma_1^2[1^n] = 0$$

$$\sigma_2^2[1^n] = 0 + \sigma_1^2\left[1\left(\frac{n-1}{n}\right);\ \infty\left(\frac{1}{n}\right)\right] = 0 + \frac{n}{(n-1)^2}$$

$$\sigma_3^2[1^n] = 0 + \frac{n}{(n-1)^2} + \sigma_1^2\left[1\left(\frac{n-2}{n}\right),\ \infty\left(\frac{2}{n}\right)\right]$$

(4.2)
$$= 0 + \frac{n}{(n-1)^2} + \frac{2n}{(n-2)^2}$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\sigma_s^2[1^n] = 0 + \frac{n}{(n-1)^2} + \frac{2n}{(n-2)^2}$$

$$+ \cdots + \frac{(s-1)n}{(n-s+1)^2} = n\sum_{i=1}^{s-1}\frac{i}{(n-i)^2}.$$

The solution for this problem for $s = n$ is given in Uspensky [9], but a straightforward solution requires a great deal of formal manipulation. The step-by-step procedure used here is somewhat indicative of the methods to be used in the succeeding portions of this paper.

4.2. *Mean and variance of the number of trials required to obtain either k_1 balls in the first box, or k_2 in the second, $\cdots$, or k_{n-1} in the $(n-1)$st, the probability associated with the nth box being non-zero.* The mean number of trials in this

particular problem is represented by $E_1[k_1(p_1), \cdots, k_{n-1}(p_{n-1}), \infty \ (p_n)]$. The formal expression for this quantity is

$$
(4.3) \quad \sum_{i=1}^{n-1} \sum_{j=k_i}^{\infty} j \frac{(j-1)!}{(k_i-1)!\,(j-k_i)!} p_i^{k_i}
$$

$$
\times \sum \frac{(j-k_i)!}{r_1!\cdots r_{i-1}!\,r_{i+1}!\cdots r_n!} p_1^{r_1}\cdots p_{i-1}^{r_{i-1}} p_{i+1}^{r_{i+1}}\cdots p_n^{r_n},
$$

where the third sum is taken over all values of the r's such that

$$
r_1 + \cdots + r_{i-1} + r_{i+1} + \cdots + r_n = j - k_i
$$

and

$$
r_1 < k_1, \cdots, r_{i-1} < k_{i-1}, r_{i+1} < k_{i+1}, \cdots, r_{n-1} < k_{n-1}.
$$

This expression can be reduced by one dimension by the application of some of the results for two boxes. Consider for the moment only those balls going into the first $(n-1)$ boxes. Then the number of balls (conditional) which is necessary to obtain either k_1 in the first box, or k_2 in the second, $\cdots$, or k_{n-1} in the $(n-1)$st box is a random variable X which takes on values

$$
k_1, k_1 + 1, \cdots, k_1 + k_2 + \cdots + k_{n-1} - (n-2)
$$

with corresponding probabilities π_j, where with no loss of generality it is assumed that $k_1 \le k_2 \le \cdots \le k_{n-1}$. π_j is given by a sum of $(n-1)$ multinomial expressions, the probability associated with the ith box now being $p_i \Big/ \Big(\sum_{i=1}^{n-1} p_i\Big)$, which will be designated by p'_i.

Under these circumstances it is apparent that

$$
(4.4) \quad E_1[k_1(p_1), \cdots, k_{n-1}(p_{n-1}), \infty \ (p_n)] = \sum_j \pi_j E_1[x_j(p_1 + \cdots + p_{n-1}), \infty \ (p_n)].
$$

However, (3.10) can be applied to each term in (4.4), leading to

$$
(4.5) \quad \frac{1}{(p_1 + p_2 + \cdots + p_{n-1})} \sum_j \pi_j x_j.
$$

Now from the definition of π_j and x_j we have

$$
(4.6) \quad
\begin{aligned}
&E_1[k_1(p_1), \cdots, k_{n-1}(p_{n-1}), \infty \ (p_n)] \\
&= \frac{1}{(p_1 + p_2 + \cdots + p_{n-1})} E_1[k_1(p'_1), k_2(p'_2), \cdots, k_{n-1}(p'_{n-1})].
\end{aligned}
$$

Similarly, the application of (3.11) gives the result that

$$
(4.7) \quad
\begin{aligned}
&\sigma_1^2[k_1(p_1), \cdots, k_{n-1}(p_{n-1}), \infty \ (p_n)] \\
&= \frac{p_n}{(p_1 + p_2 + \cdots + p_{n-1})^2} E_1[k_1(p'_1), \cdots, k_{n-1}(p'_{n-1})].
\end{aligned}
$$

These results are of immediate importance for two reasons:

1. They indicate that by combining boxes and introducing a new random variable, certain problems can be simplified. This statement will be expanded and the principle applied repeatedly in the later portions of this paper.

2. With respect to the section on two boxes, they mean that the restriction $p_1 + p_2 = 1$ is not necessary for the solution of the problems. One can always assume that $p_3(= 1 - p_1 - p_2)$ refers to a box which receives balls but which otherwise has no effect on the outcome of an experiment. In this paper it has been convenient to refer to such a box as having an infinite capacity.

4.3. *The mean value and variance of the number of trials required in a two box problem when one or both of the constants k_1 and k_2 are replaced by random variables.* The discussion in 4.2 has indicated that the idea of associating a random variable with a box instead of a single integer may sometimes lead to simplification. Here this procedure will be treated in more detail. Consider $E_1[k_1(p_1),\, k_2(p_2)]$ and assume that k_1 is replaced by a random variable X which can take on values $x_1, x_2, \cdots, x_t$ with corresponding probabilities $\pi_1, \cdots, \pi_i, \cdots, \pi_t$. Under these circumstances $E_1[\ ]$ itself becomes the random variable $E_1[X(p_1), k_2(p_2)]$, taking on values $E_1[x_i(p_1), k_2(p_2)]$, $(i = 1, 2, \cdots, t)$, with corresponding probabilities π_i. The mean value of this new random variable can be formally written down as

$$(4.8) \qquad E(E_1[X(p_1), k_2(p_2)]) = \sum_{i=1}^{t} \pi_i E_1[x_i(p_1), k_2(p_2)].$$

This expression can always be calculated from the probabilities π_i and (3.8) or from the curves given in section 6. However, in the applications which will arise later in this paper, this computation would be very time consuming. Instead, an approximation to (4.8) will now be derived which will prove to yield very good results, and which can be obtained by a simple reading on the above mentioned curves.

If X is regarded as a continuous variable, then $E_1[X(p_1), k_2(p_2)]$ is a continuous function of X, and, in fact, can be represented by a single curve similar to those appearing in section 6. Moreover, as is apparent from (3.8), repeated differentiation of $E_1[X(p_1), k_2(p_2)]$ yields continuous derivatives. Consequently, $E_1[X(p_1), k_2(p_2)]$ can be expanded in Taylor series about a, where $a = \sum_{i=1}^{t} \pi_i x_i$.
This procedure gives

$$(4.9) \qquad E(E_1[X(p_1), k_2(p_2)]) = \sum_{i=1}^{t} \pi_i \sum_{j=0}^{\infty} \frac{(x_i - a)^j}{j!} E_1^j[a(p_1), k_2(p_2)],$$

where $E_1^j[a(p_1), k_2(p_2)]$ represents the jth derivative of $E_1[X(p_1), k_2(p_2)]$ with respect to X evaluated at a. Interchanging the order of summation one obtains

$$(4.10) \qquad \sum_{j=0}^{\infty} \frac{E_1^j[a(p_1), k_2(p_2)]}{j!} \sum_{i=1}^{t} \pi_i(x_i - a)^j.$$

The final result then becomes

$$(4.11) \qquad E(E_1[X(p_1),\, k_2(p_2)]) \;=\; \sum_{j=0}^{\infty} \frac{E_1^j[a(p_1),\, k_2(p_2)]}{j!}\, \mu_j,$$

where μ_j is the jth moment of X about its mean, a. Thus to a first approximation

$$(4.12) \qquad E(E_1[X(p_1),\, k_2(p_2)]) \;\simeq\; E_1[a(p_1),\, k_2(p_2)].$$

It is of interest to note that if $E_1[X(p_1),\, k_2(p_2)]$ is linear in X then (4.12) is an exact expression since all derivatives except the first are zero. Furthermore, if $E_1[X(p_1),\, k_2(p_2)]$ is of the second degree in X, then only the second non-zero term on the right hand side of (4.11) needs to be added to (4.12) in order to make it exact. The former of these is the relation which gave an exact solution in 4.2.

It is important to realize that this analysis for $E(E_1[X(p_1),\, k_2(p_2)])$ can be immediately applied to $E(E_2[X(p_1),\, k_2(p_2)])$. For, by the use of (3.14) and (4.8), one obtains

$$(4.13) \qquad E(E_2[X(p_1),\, k_2(p_2)]) \;=\; \frac{a}{p_1} + \frac{k_2}{p_2} - E(E_1[X(p_1),\, k_2(p_2)]).$$

The same analysis can be applied to $F_{h,1}[\;\;]$ and the general result obtained that

$$(4.14) \qquad E(F_{h,1}[X(p_1),\, k_2(p_2)]) \;\simeq\; F_{h,1}[a(p_1),\, k_2(p_2)].$$

This immediately allows one to approximate the variance in the obvious manner.

It is of interest to consider briefly the situation when both k_1 and k_2 are replaced by random variables. Let k_1 be replaced by X_1 taking on values x_{11}, x_{12}, $\cdots$, x_{1t} with probabilities π_{11}, π_{12}, $\cdots$, π_{1t} and k_2 be replaced by X_2 taking on values x_{21}, x_{22}, $\cdots$, x_{2s} with probabilities π_{21}, π_{22}, $\cdots$, π_{2s}. Then

$$(4.15) \qquad E(E_1[X_1(p_1),\, X_2(p_2)]) \;=\; \sum_{i,j} \pi_{1i}\pi_{2j} E_1[x_{1i}(p_1),\, x_{2i}(p_2)],$$

where $i = 1, 2, \cdots, t$ and $j = 1, 2, \cdots, s$. Again applying Taylor series and expanding about $a = \sum_i \pi_{1i}x_{1i}$ and $b = \sum_j \pi_{2j}x_{2j}$, the result is obtained that

$$(4.16) \qquad E(E_1[X_1(p_1),\, X_2(p_2)]) \;=\; \sum_{u,v=0}^{\infty} \frac{E_1^{uv}[a(p_1),\, b(p_2)]}{u!\,v!}\, \mu_{1u}\mu_{2v},$$

where $E_1^{uv}[a(p_1),\, b(p_2)]$ is the uth partial derivative with respect to X_1 and the vth partial derivative with respect to X_2 of $E_1[X_1(p_1),\, X_2(p_2)]$ evaluated at $X_1 = a$, $X_2 = b$. This gives the approximate formula

$$(4.17) \qquad E(E_1[X_1(p_1),\, X_2(p_2)]) \;\simeq\; E_1[a(p_1),\, b(p_2)].$$

4.4. *Mean and variance of the number of trials required to obtain either* (at least) k_1 *balls in the first box, or* (at least) k_2 *balls in the second box,* $\cdots$, *or* (and at least) k_n *balls in the nth box.* In accordance with previous notation the mean number of trials required is given by $E_1[k_1(p_1), k_2(p_2), \cdots, k_n(p_n)]$. The exact value of this quantity can be written down and it would be a complicated multinomial expression. The evaluation of such an expression would be extremely difficult, if not impossible, especially for large values of k_1, k_2, $\cdots$, k_n. In order to obtain an approximation to $E_1[\ \ ]$, repeated applications of (4.12) can be made and the resulting expression can be evaluated by means of the curves in section 6.

For convenience, consider $E_1[k_1(p_1), k_2(p_2), k_3(p_3), k_4(p_4)]$. The general result will then be apparent. Assume that the first three boxes form a single unit with probability $(p_1 + p_2 + p_3)$. Then the number of balls required to obtain either k_1 in the first, k_2 in the second or k_3 in the third, if all balls are going in these three boxes, is a random variable X. Consequently,

$$(4.18) \qquad E_1[k_1(p_1), \cdots, k_4(p_4)] = E(E_1[X(p_1 + p_2 + p_3), k_4(p_4)]).$$

Applying (4.12),

$$E_1[k_1(p_1), \cdots, k_4(p_4)] \simeq$$

$$(4.19) \qquad E_1\left[E_1\left[k_1\left(\frac{p_1}{p_1 + p_2 + p_3}\right), k_2\left(\frac{p_2}{p_1 + p_2 + p_3}\right), k_3\left(\frac{p_3}{p_1 + p_2 + p_3}\right)\right]\right.$$
$$\left. (p_1 + p_2 + p_3), k_4(p_4)\right].$$

Applying (4.12) once again the final approximation is

$$E_1[k_1(p_1), \cdots, k_4(p_4)] \simeq$$

$$(4.20) \qquad E_1\left[E_1\left[E_1\left[k_1\left(\frac{p_1}{p_1 + p_2}\right), k_2\left(\frac{p_2}{p_1 + p_2}\right)\right]\left(\frac{p_1 + p_2}{p_1 + p_2 + p_3}\right),\right.\right.$$
$$\left.\left. k_3\left(\frac{p_3}{p_1 + p_2 + p_3}\right)\right](p_1 + p_2 + p_3), k_4(p_4)\right].$$

Expression (4.20) can be translated into a course of procedure. One considers the first two boxes and computes

$$a_1 = E_1\left[k_1\left(\frac{p_1}{p_1 + p_2}\right), k_2\left(\frac{p_2}{p_1 + p_2}\right)\right].$$

It is then assumed that a_1 is a new number associated with a box with probability $(p_1 + p_2)$ and

$$a_2 = E_1\left[a_1\left(\frac{p_1 + p_2}{p_1 + p_2 + p_3}\right), k_3\left(\frac{p_3}{p_1 + p_2 + p_3}\right)\right].$$

Repeating this procedure again, one computes $a_3 = E_1[a_2(p_1 + p_2 + p_3), k_4(p_4)]$, and by (4.20) this is approximately equal to $E_1[k_1(p_1), \cdots, k_4(p_4)]$. This method of computation is seen to be completely general and one can apply it to any number of boxes. Each step consists of computing $E_1[\ ]$ for two boxes and consequently can be carried out with the curves of section 6. It is evident that the order in which the boxes are taken may have an important effect on the size of the error involved in using this step-by-step procedure. This problem will be considered in section 5.

It is of interest to note that one can also obtain another approximation for $E_1[k_1(p_1), k_2(p_2), k_3(p_3), k_4(p_4)]$. Suppose that the first two boxes are considered as one unit and the second two boxes as another unit. Then the number of balls which must fall in the first two boxes in order to obtain either k_1 in the first box or k_2 in the second is a random variable X_1. Similarly a random variable X_2 can be associated with the last two boxes. Accordingly

$$(4.21) \qquad E_1[k_1(p_1), \cdots, k_4(p_4)] = E(E_1[X_1(p_1 + p_2), X_2(p_3 + p_4)]).$$

By use of (4.17), (4.21) can be written as

$$
\begin{aligned}
(4.22) \quad E_1[k_1(p_1), \cdots, k_4(p_4)] &\simeq E_1\left[E_1\left[k_1\left(\frac{p_1}{p_1 + p_2}\right), k_2\left(\frac{p_2}{p_1 + p_2}\right) \right] (p_1 + p_2), \right. \\
&\left. E_1\left[k_3\left(\frac{p_3}{p_3 + p_4}\right), k_4\left(\frac{p_4}{p_3 + p_4}\right) \right] (p_3 + p_4) \right].
\end{aligned}
$$

This same analysis applies directly to the factorial moments. In particular

$$
\begin{aligned}
(4.23) \quad F_{2,1}[k_1(p_1), \cdots, k_4(p_4)] &\simeq F_{2,1} \\
&\left[E_1\left[E_1\left[k_1\left(\frac{p_1}{p_1 + p_2}\right), k_2\left(\frac{p_2}{p_1 + p_2}\right) \right] \left(\frac{p_1 + p_2}{p_1 + p_2 + p_3}\right), \right.\right. \\
&\left.\left. k_3\left(\frac{p_3}{p_1 + p_2 + p_3}\right) \right] (p_1 + p_2 + p_3), k_4(p_4) \right].
\end{aligned}
$$

From (4.20) and (4.23) an approximate value for $\sigma_1^2[k_1(p_1), k_2(p_2), k_3(p_3), k_4(p_4)]$ can be obtained. This procedure is also perfectly general and so an estimate of $\sigma_1^2[\ ]$ can be obtained for any number of boxes.

This same method can be immediately applied to the approximation of $E_n[k_1(p_1), \cdots, k_n(p_n)]$. One simply considers the boxes two at a time, computing $E_2[\ ]$ at each stage instead of $E_1[\ ]$.

4.5. *Solution for* $E_s[k^n]$ *and* $E_s[k_1^{n-1}, k_2]$. When s is different from 1 or n, the complexities of the problem force one into the consideration of only the quantities given in the title of this subsection. The corresponding problem for three boxes, namely $E_2[k_1(p_1), k_2(p_2), k_3(p_3)]$, has been treated for general k_i and p_i by McCarthy [5]. However, the resulting expression is so complicated that it will not be given here.

The process to be used consists of reducing the subscript s by a series of steps

until the subscript 2 is reached. This expression can then be evaluated by the use of the curves or by simple computation. For the sake of convenience, the case $E_3[k^4]$ will be considered in detail. It will then be possible to write down the expression for general s and n.

As a starting point, look upon the first three boxes as a single unit. Then there is a definite probability π_i that one of these boxes will have k balls in it for the first time on the x_ith throw into these three boxes and that the other two boxes of the unit will each have less than k balls. Then if one of the other of the three boxes has u balls ($u < k$) the third box will have ($x_i - k - u$) balls, ($x_i - k - u < k$). Meanwhile the fourth box will also have been receiving balls, and the number in it at this time will be denoted by j, ($j = 0, 1, 2, \cdots, \infty$). For each x_i there is a probability associated with u, namely $P(u \mid x_i)$, and another probability associated with j, $P(j \mid x_i)$. For the moment, consider that box 1 has received k balls, box 2 the ($x_i - k - u$) balls, box 3 the u balls and box 4 the j balls. This numbering is of course immaterial since the situation is symmetric with respect to the first three boxes.

Now if $j \geq k$, either ($2k + u - x_i$) balls will be required in the second box or ($k - u$) balls in the third box in order to obtain three properly occupied boxes. On the other hand, if $j < k$, the specified number will be required in any two of boxes two, three and four. Consequently, with this conditional description of the situation, the required number of balls necessary to obtain three out of the four boxes occupied in the proper manner is

$$(4.24) \qquad x_i + j + E_2[(2k + u - x_i), (k - u), (k - j)],$$

where ($k - j$) will be taken as zero if j is greater than or equal to k. From this description, it is evident that the desired mean value may be obtained by summing (4.24) over all possible values of x_i, j and u. Therefore

$$E_3[k^4] = \sum_i \pi_i \left\{ x_i + \sum_{j=0}^{\infty} P(j \mid x_i) \right.$$

$$(4.25)$$

$$\left. \left(j + \sum_u P(u \mid x_i) E_2[(2k + u - x_i), (k - u), k - j)] \right) \right\}.$$

It is to be noticed that the probabilities inside the $E_2[\ ]$ in (4.24) and (4.25) do not add to one but only to 3/4. This can be easily remedied by the application of a formula similar to (4.6) and the result is obtained that

$$E_3[k^4] = \sum_i \pi_i \left\{ x_i + \sum_{j=0}^{\infty} P(j \mid x_i) \right.$$

$$(4.26)$$

$$\left. \cdot \left(j + 4/3 \sum_u P(u \mid x_i) E_2[(2k + u - x_i), (k - u), (k - j)] \right) \right\},$$

where each probability inside $E_2[\ ]$ is now 1/3.

By simple considerations

(4.27)
$$P(u \mid x_i) = \frac{\dfrac{(x_i - k)!}{u!(x_i - k - u)!} (\tfrac{1}{4})^u (\tfrac{1}{4})^{x_i - k - u}}{\sum_u \dfrac{(x_i - k)!}{u!(x_i - k - u)!} (\tfrac{1}{4})^u (\tfrac{1}{4})^{x_i - k - u}},$$

where u and $(x_i - k - u)$ are both less than k, and

(4.28)
$$P(j \mid x_i) = \frac{(x_i + j - 1)!}{(x_i - 1)! j!} (\tfrac{2}{4})^{x_i} (\tfrac{1}{4})^j.$$

From (4.27) and (4.28)

(4.29)
$$\sum_j j P(j \mid x_i) = x_i/3,$$

and

(4.30)
$$\sum_u u P(u \mid x_i) = \frac{x_i - k}{2}.$$

(4.25) can be written as

(4.31)
$$E_3[k^4] = \sum_i \pi_i x_i + \sum_i \pi_i \sum_j j P(j \mid x_i)$$
$$+ \frac{4}{3} \sum_i \pi_i \sum_j P(j \mid x_i) \sum_u P(u \mid x_i) E_2[(2k + u - x_i), (k - u), (k - j)].$$

Finally, making use of (4.29), (4.30), the definition of x_i and π_i and the procedure of replacing random variables inside an $E_2[\ ]$ by their mean values,

(4.32)
$$E_3[k^4] \simeq \frac{4}{3} \left\{ E_1[k^3] + E_2 \left[\left(\tfrac{3}{2}k - \frac{E_1[k^3]}{2} \right), \left(\tfrac{3}{2}k - \frac{E_1[k^3]}{2} \right), \left(k - \frac{E_1[k^3]}{3} \right) \right] \right\},$$

and this in turn can be written as

(4.33)
$$E_3[k^4] \simeq \frac{4}{3} \left\{ E_1[k^3] + E_2 \left[\left(\tfrac{3}{2}k - \frac{E_1[k^3]}{2} \right)^2, \left(k - \frac{E_1[k^3]}{3} \right) \right] \right\}.$$

This method of analysis which has just been applied to $E_3[k^4]$ can be used equally well for $E_s[k^n]$. Here one simply considers the first $(n - 1)$ boxes and proceeds as above. The final result is immediately apparent, namely that

(4.34)
$$E_s[k^n] \simeq \frac{n}{n-1} \left\{ E_1[k^{n-1}] + \right.$$
$$\left. E_{s-1} \left[\left(\frac{n-1}{n-2} k - \frac{E_1[k^{n-1}]}{n-2} \right)^{n-2}, \left(k - \frac{E_1[k^{n-1}]}{n-1} \right) \right] \right\}.$$

It will be noticed that in reducing (4.34) further it will be necessary to consider expressions of the form $E_s[k_1^{n-1}, k_2]$. However, it will be seen from the foregoing

analysis that no use was made of the fact that the integers attached to the first $(n - 1)$ boxes were the same.   Accordingly,

(4.35)
$$E_s[k_1^{n-1}, k_2] \simeq \frac{n}{n-1} \left\{ E_1[k_1^{n-1}] + \right.$$
$$\left. E_{s-1}\left[\left(\frac{n-1}{n-2} k_1 - \frac{E_1[k_1^{n-1}]}{n-2} \right)^{n-2}, \left(k_2 - \frac{E_1[k_1^{n-1}]}{n-1} \right) \right] \right\}.$$

Now, by the use of (4.34) and (4.35), it is possible to reduce s as much as may be desired.

5. Some considerations concerning the error of the approximations.

5.1. *Preliminary remarks.*   This discussion of the errors of the approximations given in the preceding sections has been left until now so that a broad perspective might be gained, and the errors seen in relationship to one another.   Such an arrangement is advantageous in this instance since both the analytical and computational results bearing on the subject are scanty, and consequently, any intelligent leads which their inter-relationships can give are most helpful.

The difficulty involved in obtaining exact values for the various quantities considered in this paper has been pointed out quite frequently, and the approximations have been devised to overcome this very difficulty.   The same complexity which prevents the computation of many exact values also prevents any effective analytic approach to the problem of evaluating the errors.   For these reasons the author has been unable to carry through any general analytic treatment of the errors of the approximations.   However, because the intelligent use of approximations requires some knowledge of their accuracy, certain isolated cases have been investigated by a combination of computational, graphical and analytic methods.   These investigations are detailed in the remainder of this section, and conjectures concerning the general behavior of the errors are made whenever possible.   As has been stated earlier, no consideration will be given to the approximation formulae for the variance.

5.2. *Errors of the approximations for $E_1[k_1(p_1), \cdots, k_n(p_n)]$ and*
$$E_n[k_1(p_1), \cdots, k_n(p_n)].$$

Taking n equal to 3, we have from (4.11) that

(5.1)
$$| E_1[k_1(p_1), k_2(p_2), k_3(p_o)] - E_1[a(p_1 + p_2), k_3(p_3)] |$$
$$\leq \tfrac{1}{2} \sigma_1^2 \left[k_1 \left(\frac{p_1}{p_1 + p_2} \right), k_2 \left(\frac{p_2}{p_1 + p_2} \right) \right]$$
$$\cdot \operatorname{Max} | E_1^2[X(p_1 + p_2), k_3(p_3)] |,$$

where $\operatorname{Max} | E_1^2[X(p_1 + p_2), k_3(p_3)] |$ is the maximum absolute value of the second derivative of $E_1[X(p_1 + p_2), k_3(p_3)]$ with respect to X, and a is equal to $E_1[k_1(p_1/(p_1 + p_2)), k_2(p_2/(p_1 + p_2))]$.   Now an examination of the curves

given in section 6 indicates that, for fixed p_3 and k_3, the maximum curvature of $E_1[X(p_1 + p_2), k_3(p_3)]$, considered as a function of X, is a monotone decreasing function of k_3. Since this curvature is negative, this geometric observation is equivalent to

$$\text{Max} \mid E_1^2[X(p_1 + p_2), (k_3 + 1)(p_3)] \mid$$

(5.2)

$$\leq \text{Max} \mid E_1^2[X(p_1 + p_2), k_3(p_3)] \mid,$$

although it is not necessarily true that

$$\mid E_1^2[x_1(p_1 + p_2), (k_3 + 1)(p_3)] \mid \leq \mid E_1^2[x_1(p_1 + p_2), k_3(p_3)] \mid.$$

Moreover,

(5.3) $$E_1[k_1(p_1), k_2(p_2), k_3(p_3)] \leq E_1[k_1(p_1), k_2(p_2), (k_3 + 1)(p_3)].$$

From (5.1), (5.2) and (5.3) one readily obtains that the absolute value of the percentage error of the approximation to $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$ is bounded by a function, say $U_1[k_1(p_1), k_2(p_2), k_3(p_3)]$, which is a monotone decreasing function of k_3 as k_3 increases. It should be noticed that the results of 4.2 have already shown not only that this upper bound for the percentage error approaches zero as k_3 becomes infinite, but also that the absolute difference between the true and approximate values approach zero as k_3 becomes infinite.

Computation of $U_1[k_1(p_1), k_2(p_2), k_3(p_3)]$ is very time consuming because of the difficulty in obtaining $\text{Max} \mid E_1^2[X(p_1 + p_2), k_3(p_3)] \mid$, and because the direct computation of $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$ is laborious when any of k_1, k_2 and k_3 are much larger than 2 or 3. In order to surmount these difficulties and still give some indication of the behavior of $U_1[k_1(p_1), k_2(p_2), k_3(p_3)]$, the following expedients were adopted:

1. The values of k_1, k_2 and k_3 were each fixed at 5,

2. $\text{Max} \mid E_1^2[X(p_1 + p_2), k_3(p_3)] \mid$ was obtained by graphical means, namely drawing the slopes of the appropriate curve in section 6, graphing these slopes and then taking off the maximum slopes of these curves.

3. $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$ was replaced by its approximation,

$$E_1[a(p_1 + p_2), k_3(p_3)],$$

in the computation of the percentage error. This new bound will be denoted by $U_1^*[k_1(p_1), k_2(p_2), k_3(p_3)]$.

4. Carefully chosen values of $U_1^*[k_1(p_1), k_2(p_2), k_3(p_3)]$ were plotted on triangular coordinates, and contour lines interpolated and extrapolated to cover in large part the range of p_1, p_2 and p_3.

The use of the third of the above listed assumptions is no detriment to the usefulness of the results since

$$\frac{E_{1a}[\,] - E_1[\,]}{E_1[\,]} = \frac{\dfrac{E_{1a}[\,] - E_1[\,]}{E_{1a}[\,]}}{1 - \dfrac{E_{1a}[\,] - E_1[\,]}{E_{1a}[\,]}} \leq \frac{U_1^*[k_1(p_1), k_2(p_2), k_3(p_3)]}{100 - U_1^*[k_1(p_1), k_2(p_2), k_3(p_3)]},$$

where $E_{1a}[\quad] = E_1[a(p_1 + p_2), k_3(p_3)]$ and $E_1[\quad] = E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$. Since $U_1^*[\quad]$ is a monotone decrease function of k_3, this new bound on the percentage error is also monotone decreasing for increasing k_3. Absolute values were not required in this derivation since $E_{1a}[\quad]$ is always greater than or equal to $E_1[\quad]$, as is apparent from (5.1) and an examination of the curves of section 6. The contours of $U_1^*[5(p_1), 5(p_2), 5(p_3)]$ are shown in Fig. 1. The interpretation of this figure is very straightforward. For example, for $p_3 \leqslant .5$, the value
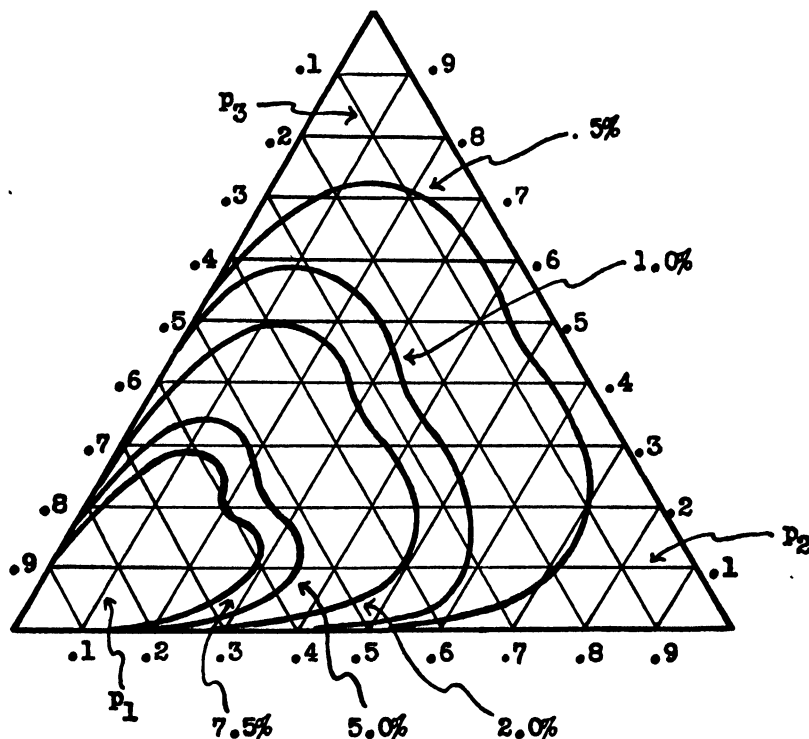
FIG. 1. CONTOURS OF $U_1^*[5(p_1), 5(p_2), 5(p_3)]$ CONSIDERED AS A FUNCTION OF p_1, p_2 AND p_3

of $U_1^*[5(p_1), 5(p_2), 5(p_3)]$ is less than 5.0%. Making use of the definition of $U_1^*[\quad]$, and especially its monotone characteristic, one can then say: the approximation for $E_1[5(p_1), 5(p_2), k_3(p_3)]$, where $k_3 \geqslant 5$, $p_3 \leqslant .50$ is in error by not more than 5.3%. Moreover, as has been already observed $E_1[a(p_1 + p_2), k_3(p_3)]$ is always greater than or equal to $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$.

It will be noticed from Fig. 1 that $U_1^*[\quad]$ is increasing steadily as p_3 approaches 1. It has been demonstrated by McCarthy [5] that this behavior of the upper bound does not mean that the percentage error itself becomes larger as p_3 ap-

proaches 1. As a matter of fact, for fixed k_1, k_2 and k_3, the percentage error approaches zero as p_3 approaches 1. However, this demonstration does not furnish any reasonable bounds with which to fill in the lower left hand corner of Fig. 1. This fact is not as serious as it may at first seem because there is nothing to prevent one from reordering the boxes. For example, consider $E_1[5(.2)$, $5(.2)$, $5(.6)]$. From Fig. 1, the error of the approximation for this quantity, namely $E_1[E_1[5(.5), 5(.5)](.4), 5(.6)]$, is not more than approximately

$$7.5/(100 - 7.5) = .8.1\%.$$

On the other hand this same figure shows that $E_1[E_1[5(.25), 5(.75)](.80), 5(.20)]$, which is also an approximation to $E_1[5(.2), 5(.2), 5(.6)]$, is in error by not more than approximately $.8\%$. Consequently one would choose the second ordering.

The procedure which has been used to obtain an upper bound on the percentage error of the approximation to $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$, k_1 and k_2 fixed and k_3 greater than or equal to that integer at which the bound is evaluated, can also be applied to $E_3[k_1(p_1), k_2(p_2), k_3(p_3)]$. All the assumptions remain the same and in this case the bounds corresponding to $U_1[\ ]$ and $U_1^*[\ ]$ are denoted by $U_3[\ ]$ and $U_3^*[\ ]$. As in the case of $U_1[\ ]$, we have

$$\frac{E_3[\ ] - E_{3b}[\ ]}{E_3[\ ]} = \frac{\dfrac{E_3[\ ] - E_{3b}[\ ]}{E_{3b}[\ ]}}{1 + \dfrac{E_3[\ ] - E_{3b}[\ ]}{E_{3b}[\ ]}} \leq \frac{U_3^*[k_1(p_1), k_2(p_2), k_3(p_3)]}{100}.$$

Here the approximation, $E_2[b(p_1 + p_2), k_3(p_3)]$, is always less than or equal to the exact value, $E_3[k_1(p_1), k_2(p_2), k_3(p_3)]$. The contours of $U_3^*[5(p_1), 5(p_2), 5(p_3)]$ are shown in Fig. 2. In using $U_3^*[5(p_1), 5(p_2), 5(p_3)]$ it is sometimes advantageous to reorder the boxes. For example, consider $E_3[5(.2), 5(.2), 5(.6)]$. Fig. 2 shows that, as an approximation, $E_2[E_2[5(.5), 5(.5)](.4), 5(.6)]$ is in error by not more than approximately 9%. However, $E_2[E_2[5(.25), 5(.75)](.80), 5(.20)]$, which is also an approximation for $E_3[5(.2), 5(.2), 5(.6)]$, is in error by not more than about 7%. There is a gain here, but it is not as great as the corresponding situation for $E_1[5(.2), 5(.2), 5(.6)]$.

As has already been stated, one may minimize the error by correctly choosing the two boxes which are to be combined first. Some discussion will be given here of a procedure for choosing these two boxes. Of course an experimental scheme may be used which makes use of the fact that the approximation to $E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$ is always an overestimate. In other words, *that grouping is used which gives rise to the smallest value of the approximation*. However, this can be replaced by a few preliminary computations.

As can be seen from (5.1), the error of the approximation depends upon two quantities, namely the variance of the two box situation obtained by combining two of the boxes, and the maximum value of the second derivative of the curve representing the function $E_1[X(p_1 + p_2), k_3(p_3)]$ over the proper range of X values. The error will be zero of $E_1[X(p_1 + p_2), k_3(p_3)]$ is either a constant or

linear in X over the range of X values in which one is interested, that is $k_1 \leq X \leq k_1 + k_2 - 1$, $k_1 \leq k_2$. If this is not possible, then one wishes to make it as near so as possible, subject to the restriction that

$$\sigma_1^2[k_1(p_1/(p_1 + p_2)), k_2(p_2/(p_1 + p_2))]$$

is not unnecessarily large.
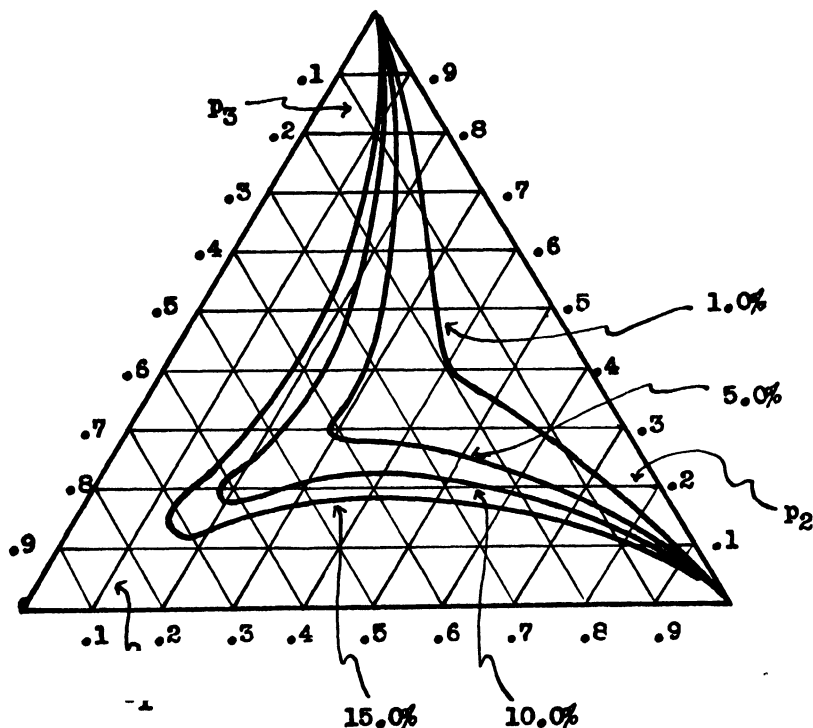
FIG. 2. CONTOURS OF $U_3^*[5(p_1), 5(p_2), 5(p_3)]$ CONSIDERED AS A FUNCTION OF p_1, p_2 AND p_3

An indication of the relationship between the boxes for both linearity and contribution to variance can be obtained from expressions (3.10) and (3.11). Thus for each box one computes k_i/p_i and $k_i(1 - p_i)/p_i^2$. Then in order to most nearly achieve linearity one orders the boxes in accordance with the increasing order of k_i/p_i and combines them in that order. If there is a tie between two or more boxes with respect to the k_i/p_i ordering, then one orders these "tied" boxes in accordance with increasing $k_i(1 - p_i)/p_i^2$.

Some computations have been carried out to illustrate these points and they are given in Table 1. The notation $((2, 4), 6)$ means that one first combines the boxes with integers 2 and 4, and then combines this result with the box with

associated integer 6. All values in this table were obtained by direct computation. No use of the curves was made.

In these three situations, one obtains the values given in Table 2.

Thus in the first case there is nothing to choose with respect to k_i/p_i, but $k_i(1 - p_i)/p_i^2$ indicates the ordering $((6, 4), 2)$. Actually the percentage error in this instance is 1.0 as compared with 1.7 and 2.4 for the other two orderings. In case two, k_i/p_i indicates the ordering $((2, 6), 4)$. Although this does not turn out to be the best ordering, Table 1 shows that the ordering in this instance makes little difference. In the last case, the indicated ordering is $((2, 4), 6)$ and the percentage error for this is zero, as opposed to 1.3 and 1.6. Since at any stage in the operation of combining boxes two at a time (4.13) holds, the

TABLE 1

Effect of Order of Combination on Error of Approximation

p_1 $1/6$ k_1	p_2 $1/3$ k_2	p_3 $1/2$ k_3	$E_1[k_1(p_1), k_2(p_2), k_3(p_3)]$	% Error of Approximation		
				Order of Combination		
				$((2, 4), 6)$	$((2, 6), 4)$	$((4, 6), 2)$
2	4	6	6.96	+1.7	+2.4	+1.0
4	6	2	3.92	+0.3	+0.5	+0.5
6	4	2	3.77	+0.0	+1.3	+1.6

TABLE 2

p_i	1/6	1/3	1/2	1/6	1/3	1/2	1/6	1/3	1/2
k_i	2	4	6	4	6	2	6	4	2
k_i/p_i	12	12	12	24	18	4	36	12	4
$k_i(1 - p_i)/p_i^2$	60	24	12	120	36	4	180	24	4

above procedure will also give the minimum error for the approximation to $E_3[k_1(p_1), k_2(p_2), k_3(p_3)]$. *Moreover, the approximation for this quantity is always an underestimate of the true value, and therefore that ordering should be taken which gives the greatest value for the approximation.*

When the error of the approximation to $E_1[k_1(p_1), \cdots, k_n(p_n)]$ and

$$E_n[k_1(p_1), \cdots, k_n(p_n)],$$

for n greater than three, is considered, it is immediately obvious that the general considerations already given in this section still apply. In addition to these considerations, there is the difficulty that errors may cumulate. However, the results already quoted for three boxes, in conjunction with those which are to be given in 5.3, indicate that this cumulation is not serious. There are two factors which eventually prevent (i.e. as more and more boxes are considered) this percentage error from becoming unduly large, and, in fact, make it approach zero. These are:

1. The value of p_3 will, in most instances, be decreasing as more and more boxes are considered (see Fig. 1), and

2. The true value is usually becoming larger and larger as more and more boxes are considered.

In order to minimize the error, the following precautions should be taken:

1. At each stage in the computation, try to avoid, as much as possible, making readings where $E_1[X(p_1' + p_2'), k_3(p_3')]$ is curving sharply. If all readings are made where the curves are nearly linear, the percentage error will be very close to zero. On the other hand, if many readings must be made where the slopes of the curves are changing most sharply, larger errors must be expected.

2. Use that ordering of the boxes which provides the minimum value for the approximation to $E_1[\ ]$ or the maximum value for the approximation to $E_n[\ ]$.

3. In order to approximate the ordering which (2) would give, compute k_i/p_i and $k_i(1 - p_i)/p_i^2$ at each stage at which two boxes are to be combined and use the rules of procedure already given for three boxes.

5.3. Error of the approximation for $E_s[k^n]$. Repeated applications of the reduction formulae (4.34) and (4.35) allow one to evaluate $E_s[k^n]$ by means of the solution for the two box case, or more explicitly, by means of the curves given in section 6. Here the error of this approximation will be discussed primarily from a computational point of view.

$E_s[1^n]$ can be treated in detail since it is possible to obtain exact values for this expression by means of (4.1). This has been done by McCarthy [5], but the details will not be repeated here because of lack of space. The results simply add more credence to the conjectures which will soon be made.

When k is taken to be larger than one, the difficulty arises that it is almost impossible to compute the exact value of $E_s[k^n]$ in a large number of cases. Consequently it was necessary to devise an experimental model to estimate these exact values so that the amount of error would be known within bounds. A set of 10,000 punched cards[1] was obtained on which were recorded 100,000 random numbers drawn from a rectangular distribution. Thus if the cards are ordered on a particular set of columns, and one reads off the digits 0–9 on another specified column, one card at a time, it is equivalent to using a table of random numbers such as those prepared by Tippett [7]. By the use of these cards, it was possible to run off on an IBM Tabulator any desired number of experiments in order to obtain an experimental distribution from which to calculate an estimate of $E_s[k^n]$ and the variance of this estimate. For example, in determining an estimate of $E_1[2^5]$ one hundred experimental trials were made, as described above, with the following results:

Number of Trials Required	Frequency
2	23
3	32
4	31
5	11
6	3

[1] These punched cards were prepared at the Mayo Clinic, Rochester, Minn., under the direction of Doctor Joseph Berkson.

From this distribution the estimate of $E_1[2^5]$ is 3.39, with a variance computed from the distribution of .011. The 95% symmetric confidence limits for the mean, computed from the Student t-distribution, are 3.17 and 3.61. Such estimates will be used in the remainder of this section. It should be pointed out that in order to prevent a prohibitive amount of machine time, it was

TABLE 3

Percentage Errors for $E_s[k^n]$

s	k	n 3	4	5
1	1	—	—	—
	2	+ .7	+ 2.2	− .3 +13.6
	5	+ 1.1	− 3.1 +5.7	+ .6 +10.7
	10			− 2.9 + 5.1
2	1	− 5.6	− .4	+ 1.3
	2	− 4.6	− 4.4 +4.4	+ .6 +10.4
	5	−4.6 +1.7	+ 3.0 +9.3	+ 7.9 +14.8
	10	−3.7 +2.1	− .3 +5.5	+ 4.3 +10.7
	15		+ 1.0 +7.2	
	20	−2.5 +2.4		
3	1	−18.2	−12.7	− 3.1
	2	− 6.3	−16.5 −7.3	− 2.9 + 6.0
	5	−9.7 −2.2	−10.7 −5.5	+ .8 + 5.8
	10			− 2.1 + 3.1
4	1		−12.0	−15.6
	2		−13.6 +6.1	−11.6 − 3.9
	5		−13.9 −7.2	− 9.9 − 4.0
	10		− 8.9 −2.6	− 6.4 − 1.2
5	1			−8.8
	2			−18.1 − 6.0
	5			−12.5 − 5.6
	10			− 8.9 − 2.9

necessary to use many of the same runs to determine values of $E_s[k^n]$ for different values of s, k and n. This means that the errors are correlated to some slight extent, but it would be extremely difficult to determine how much.

A summary of the computed percentage errors for various values of s, k and n is given in Table 3. In the instances where there are two entries, they are calculated on the basis of the 95% confidence limits for the experimental mean. These confidence limits are symmetric and were determined by using the Student t-distribution. For k equal to 2 and 5 the distribution contained 100 trials,

while for k greater than 5, the distribution were made up of approximately 50 trials.

The computations given in this table show for various values of s, k and n, the percentage error of the approximation for $E_s[k^n]$. In addition to showing the values of these percentage errors, the computations lead one to conjecture that

1. For fixed s and k, there exists an n_0 such that for $n > n_0$ the absolute value of the percentage error of the approximation for $E_s[k^n]$ is a monotone decreasing function for increasing n. It was shown by McCarthy [5] that this absolute value approaches zero as n approaches infinity for $E_s[1^n]$, and in fact, that the difference between the true and approximate values approaches zero.

2. For fixed s and n, there exists a k_0 such that for $k > k_0$, the absolute value of the percentage error of the approximation for $E_s[k^n]$ is a monotone decreasing function for increasing k.

6. Computation.

6.1. Curves to aid in the computation of $E_1[k_1(p_1), k_2(p_2)]$. In 3.1 it was shown that $E_1[k_1(p_1), k_2(p_2)]$ is equal to

$$\frac{k_1}{p_1} I_{p_1}(k_1 + 1, k_2) + \frac{k_2}{p_2} I_{p_2}(k_2 + 1, k_1),$$

where $I_x(p, q)$ is the Incomplete Beta-Function as tabled by Karl Pearson [6]. There are three principal difficulties connected with the use of these tables as they apply to the approximations of this paper. These are:

1. The tables must be available,

2. The tables give directly only values for integer or half-integer values of k_1 and k_2, and

3. Since many different values of $E_1[k_1(p_1), k_2(p_2)]$ are often required to obtain a single approximation, the computational burden would be very heavy.

In order to surmount these difficulties, it seemed advisable to prepare curves giving the values of $E_1[k_1(p_1), k_2(p_2)]$ for various values of k_1, k_2, p_1 and p_2. These curves would give values of $E_1[\ \ ]$ with sufficient accuracy for most problems not only for integer values of k_1 and k_2, but for all values over the range considered.

Such curves have been prepared by computing $E_1[k_1(p_1), k_2(p_2)]$ for integral values of k_1 and k_2 (for fixed p_1 and p_2) and then joining these points with a smooth curve. A summary of the graphs prepared is as follows:

	k_1	k_2	p_1	p_2
Fig. 3	$1, 2, \cdots, 25, \infty$	$1, 2, \cdots, 35$	.50	.50
Fig. 4	$1, 2, \cdots, 20, \infty$	$1, 2, \cdots, 35$	.40	.60
Fig. 5	$1, 2, \cdots, 15, \infty$	$1, 2, \cdots, 35$	.20	.80
Fig. 6	$1, 2, \cdots, 10, \infty$	$1, 2, \cdots, 15$	.80	.20
Fig. 7	$1, 2, \cdots, 7, \infty$	$1, 2, \cdots, 15$	.60	.40
Fig. 8	$1, 2, \cdots, 8, \infty$	$1, 2, \cdots, 15$	.50	.50
Fig. 9	$1, 2, \cdots, 6, \infty$	$1, 2, \cdots, 15$	.40	.60
Fig. 10	$1, 2, \cdots, 5, \infty$	$1, 2, \cdots, 15$	.20	.80

Figure 3

$E_1\left[K_1(50), K_2(50)\right]$,
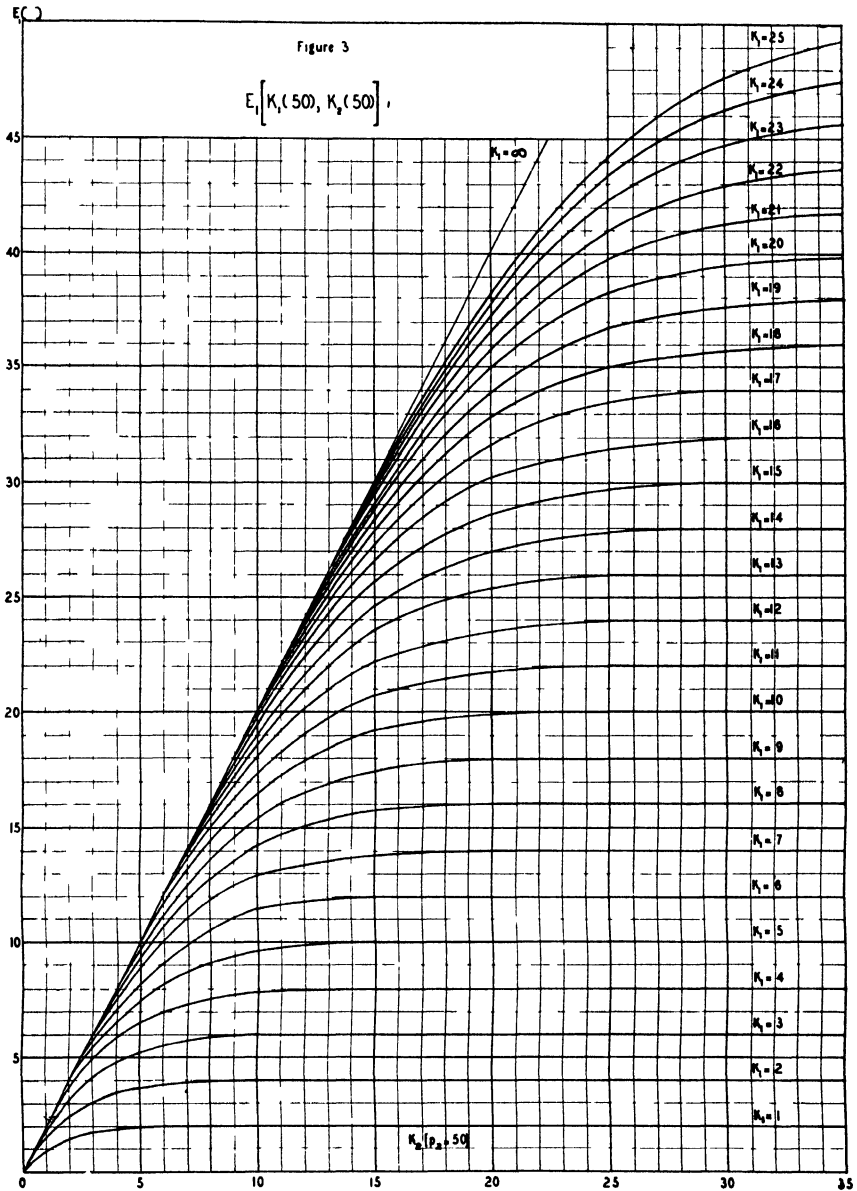
Figures 8, 9, and 10 are simply portions of figures 3, 4 and 5 drawn on an expanded scale in order to permit greater accuracy in reading the curves. Also figures 6 and 10 and figures 7 and 9 form pairs in that a member of one pair can

Figure 4

$$E_1\left[K_1(.40), K_2(.60)\right]$$
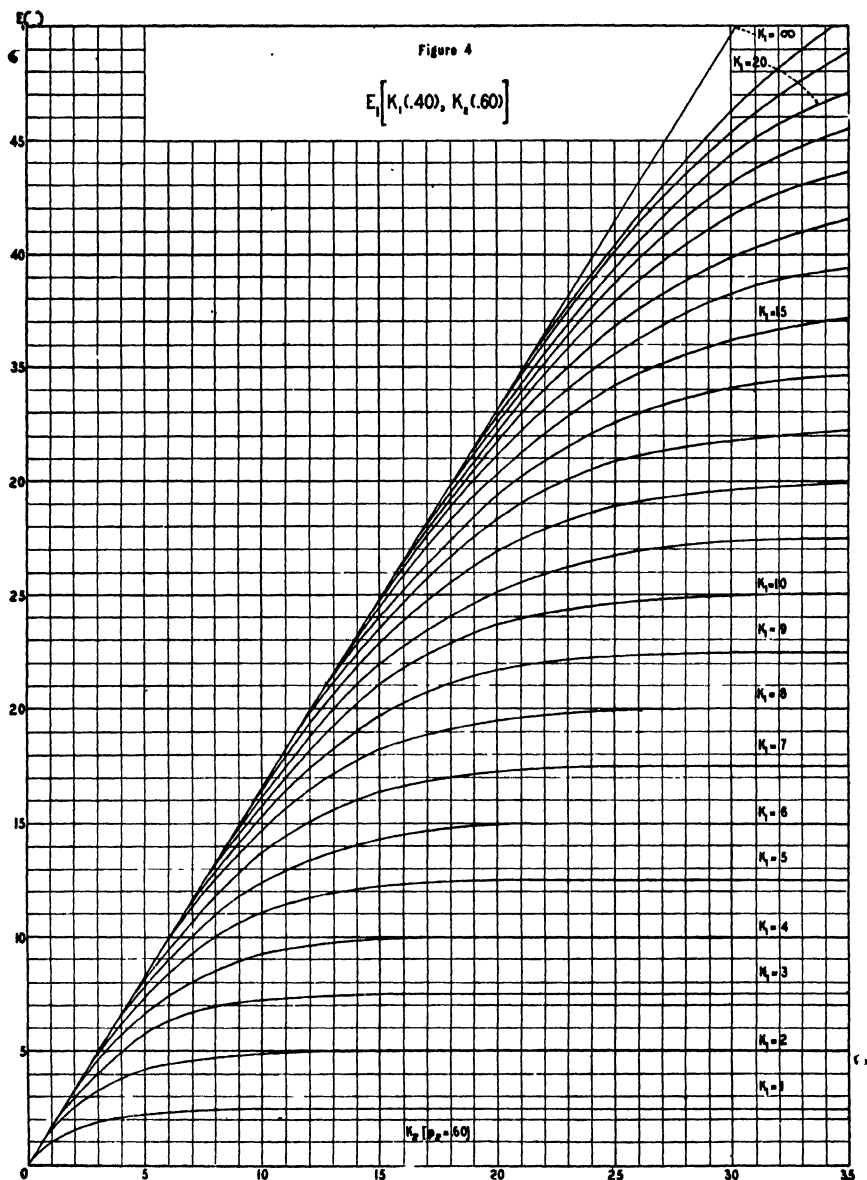
be obtained from the other member of the pair. Both members of the pair are given on the expanded scale in order to facilitate interpolation. Values of the mean for combinations of k_1 and k_2 not given directly can usually be obtained

Figure 5

$$E_1\left[K_1(.20), K_2(.80)\right]$$
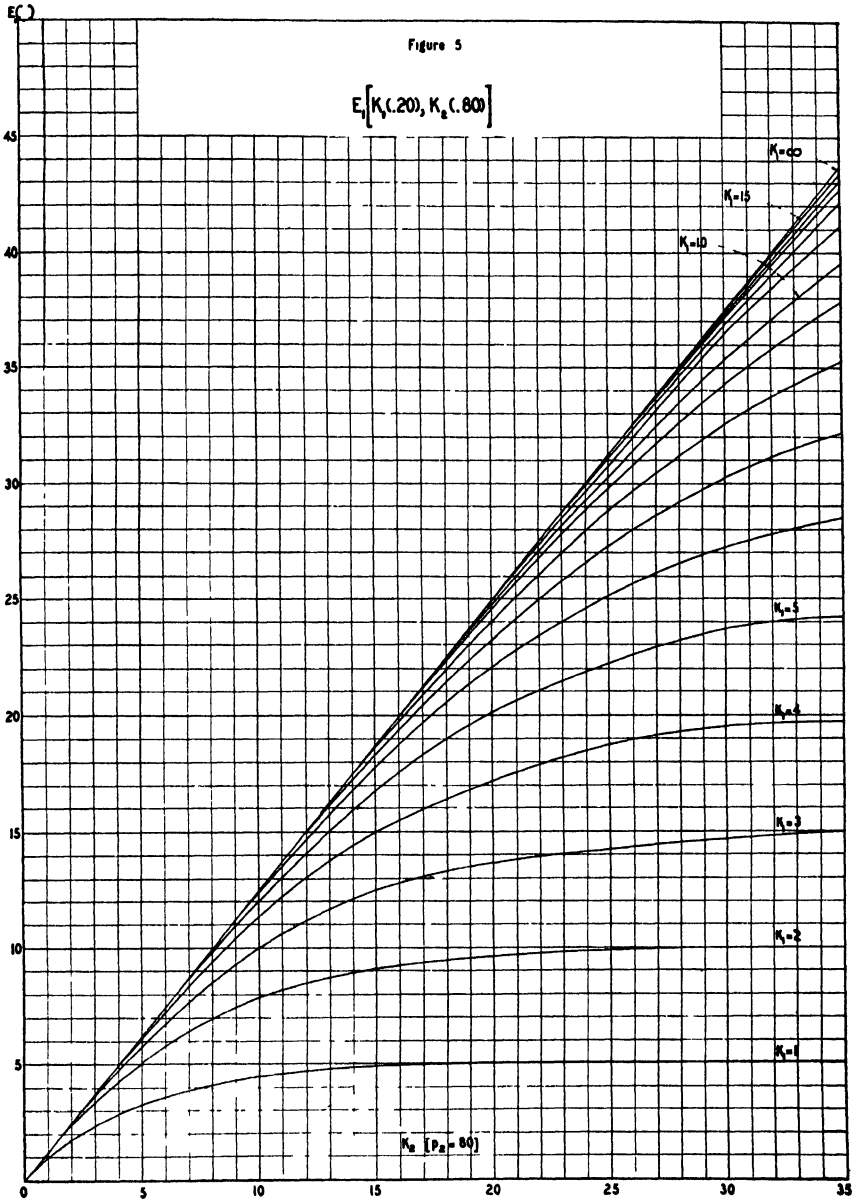
with sufficient accuracy with linear interpolation. Interpolation for p_1 and p_2 should be done graphically since in some instances linear interpolation would be extremely poor.

As an example, suppose one has two boxes with $k_1 = 2$, $k_2 = 5$, $p_1 = .40$ and $p_2 = .60$. Consulting Fig. 9, one goes along the horizontal axis to $k_2 = 5$.

FIGURE 6

$$E_1\left[K_1(.80), K_2(.20)\right]$$
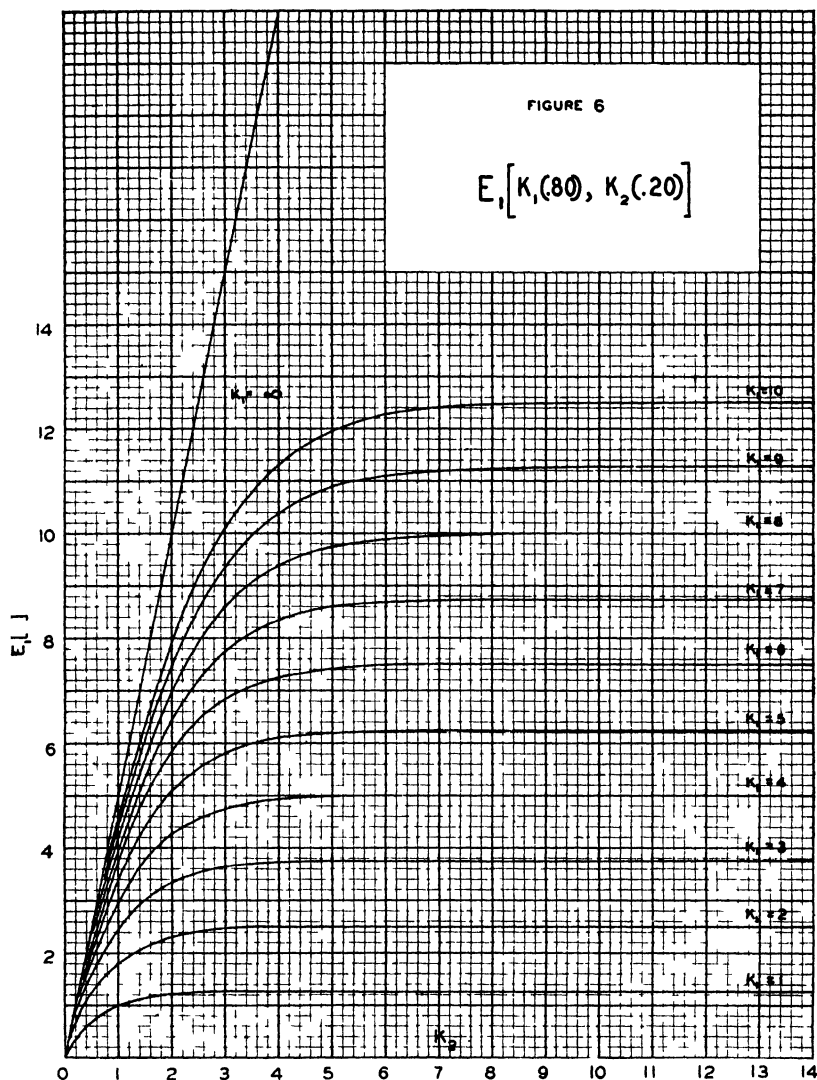
Following up the vertical line through this point to the curve $k_1 = 2$, $E_1[2(.40), 5(.60)]$ is read as 4.25.   The actually computed value to four decimals is 4.2224.

FIGURE 7

$$E_1\left[K_1(.60),\ K_2(.40)\right]$$
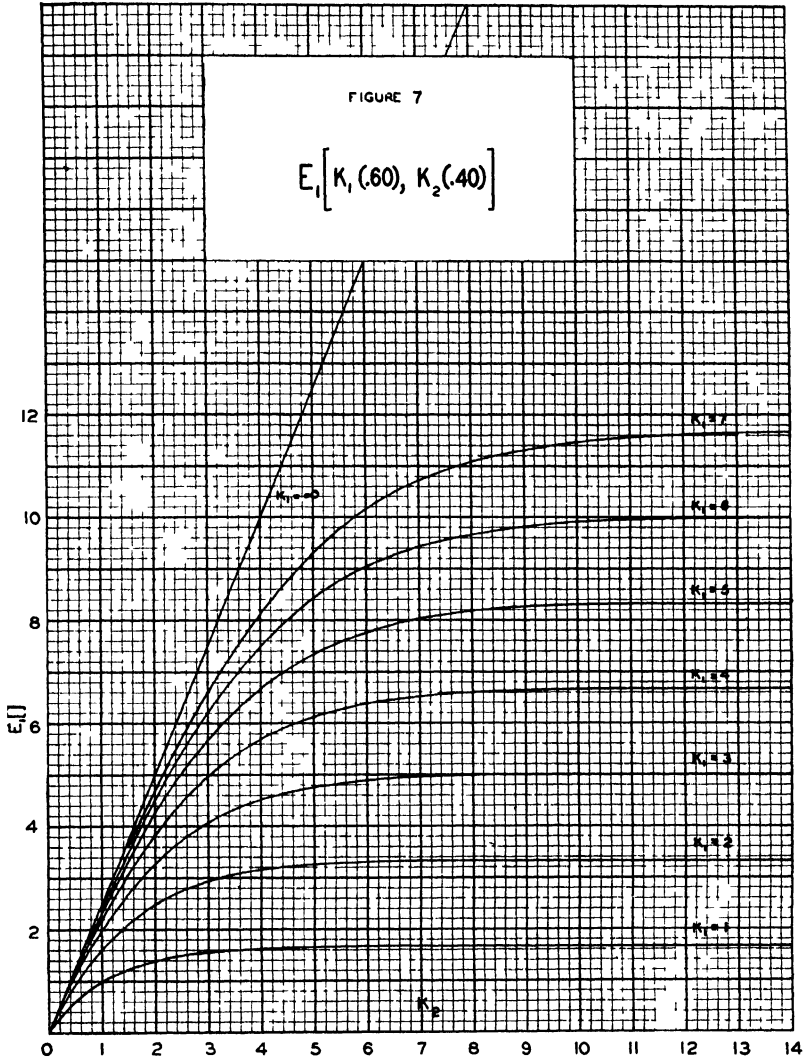
It is immediately evident that $E_2[k_1(p_1),\ k_2(p_2)]$ can also be obtained from the curves since

$$E_2[k_1(p_1),\ k_2(p_2)] = (k_1/p_1) + (k_2/p_2) - E_1[k_1(p_1),\ k_2(p_2)].$$

6.2. *Use of the curves to obtain exact values (i.e. subject only to the error of reading the curves) for $E_1[k_1(p_1),\ k_2(p_2),\ k_3(p_3)]$.* Referring back to (4.8), one obtains that

(6.1) $$E_1[k_1(p_1),\ k_2(p_2),\ k_3(p_3)] = \sum \pi_i E_1[x_i(p_1 + p_2),\ k_3(p_3)],$$

FIGURE 8

$$E_1\left[K_1(.50),\ K_2(.50)\right]$$
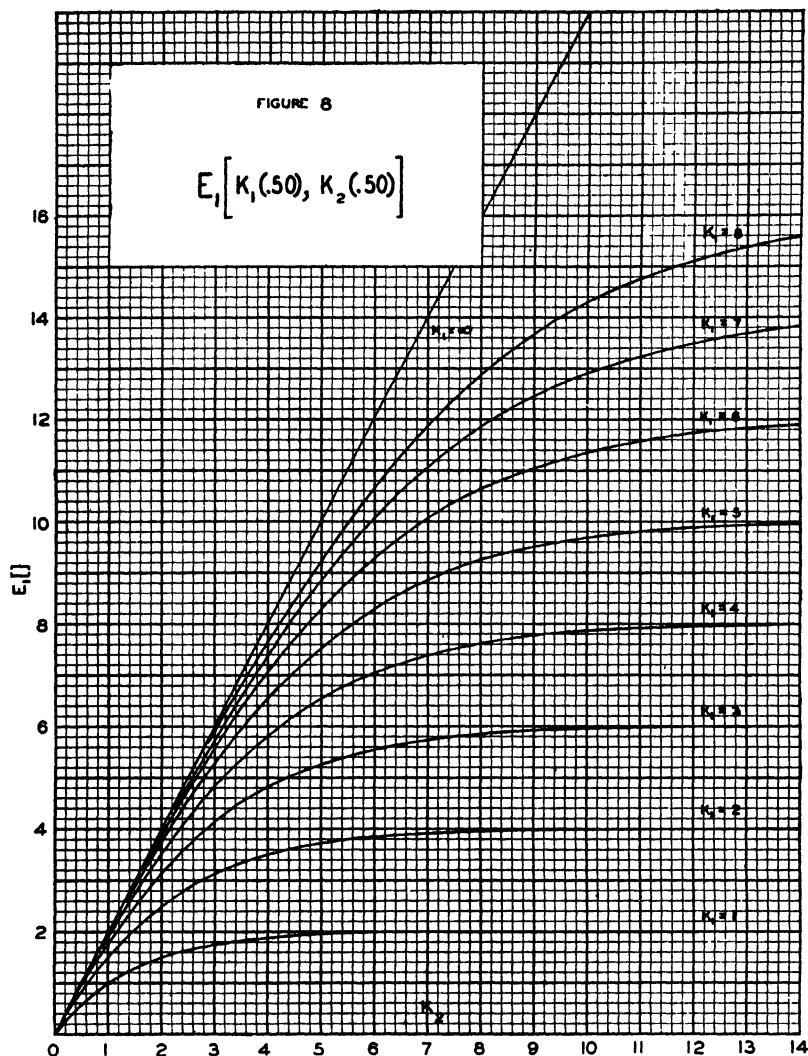
where π_i is the probability that either k_1 balls are obtained in the first box or k_2 balls are obtained in the second box on the $x_i\,th$ throw for the first time, assuming balls can go only in boxes one and two. x_i takes on values

$$k_1,\ k_1 + 1,\ \cdots,\ k_1 + k_2 - 1$$

when $k_1 \leq k_2$.   Now π_i can be easily computed and $E_1[x_i(p_1 + p_2),\ k_3(p_3)]$ can be obtained from the curves.   The only difficulty in using this procedure

FIGURE 9

$$E_1\left[K_1(.40),\ K_2(.60)\right]$$
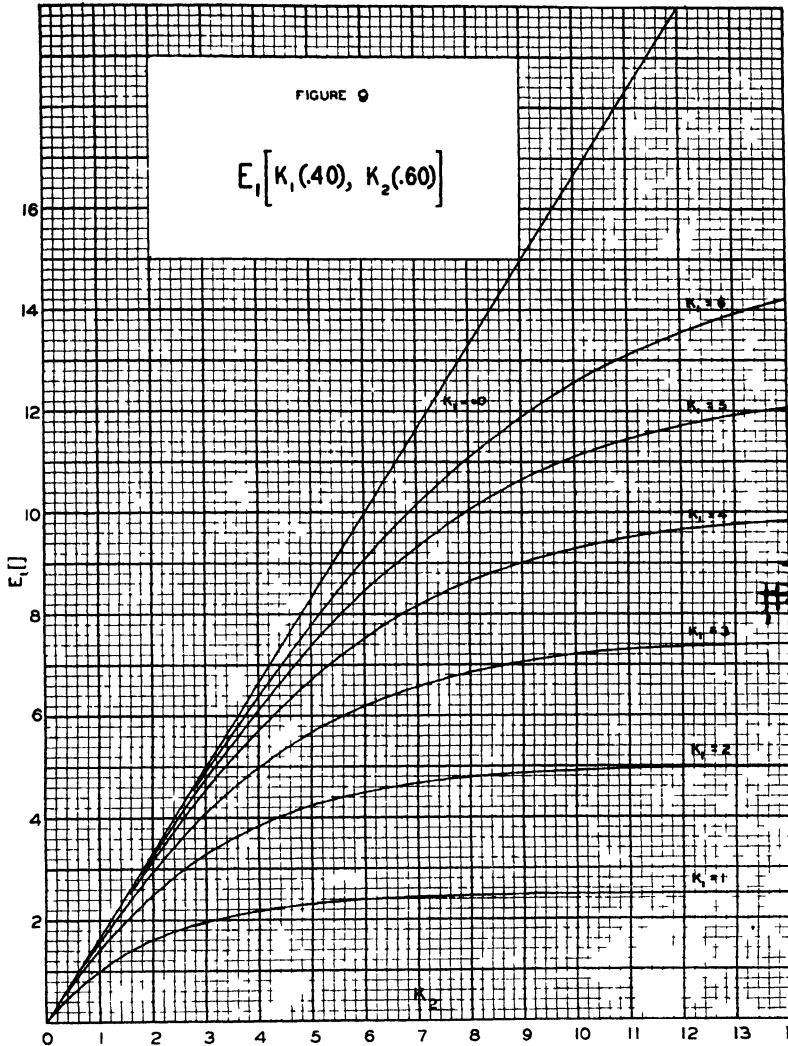
arises when the range of x_i is large. Then a large amount of computation is involved.

In order to illustrate this computation, consider $E_1[2(.1),\ 3(.1),\ 5(.8)]$. Here x_i takes on the values 2, 3 and 4.   We have $x_1 = 2,\ \pi_1 = 2/8;\ x_2 = 3,\ \pi_2 = 3/8;$ and $x_3 = 4,\ \pi_3 = 3/8$. From Fig. 6

$$E_1[2(.2),\ 5(.8)] = 5.09$$
$$E_1[3(.2),\ 5(.8)] = 5.88$$
$$E_1[4(.2),\ 5(.8)] = 6.11.$$

FIGURE 10

$$E_1\left[K_1(.20),\ K_2(.80)\right]$$
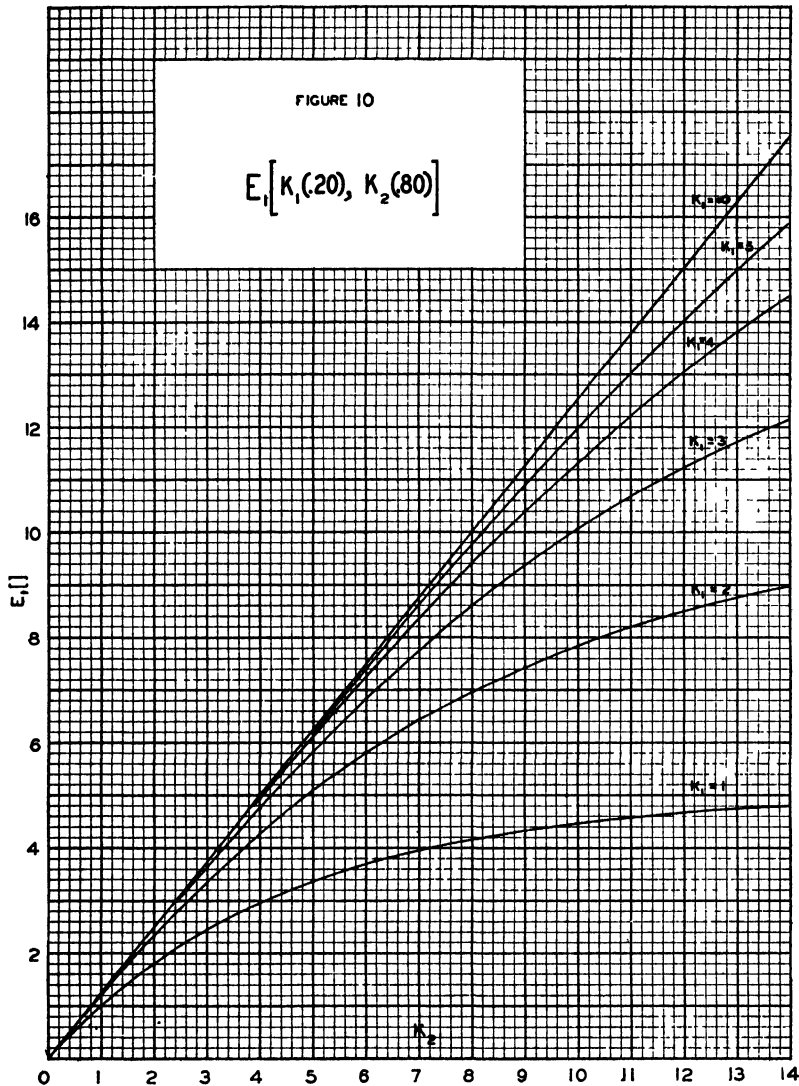
Consequently, $E_1[2(.1),\ 3(.1),\ 5(.8)]$ is equal to

$$(5.09)(2/8) + (5.88)(3/8) + (6.11)(3/8) = 5.77.$$

Using computed values for $E_1[x_i(.2),\ 5(.8)]$, $E_1[2(.1),\ 3(.1),\ 5(.8)]$ is equal to 5.75. Thus the use of the curves has only led to an error of .3%.

6.3. *Use of the curves in approximating* $E_1[k_1(p_1), \cdots, k_n(p_n)]$,

$$E_n[k_1(p_1), \cdots, k_n(p_n)]$$

and $E_s[k^n]$. In illustrating the application of the curves and the reduction formulae (4.34) and (4.35), one example will be worked through in detail. This example will provide illustrations of all the details involved in such problems. Consider $E_4[5^5]$. Applying formula (4.34)

$$(6.2) \quad E_4[5^5] \simeq 5/4 \left\{ E_1[5^4] + E_3 \left[\left(\frac{4}{3} \cdot 5 - \frac{E_1[5^4]}{3} \right)^3, \quad \left(5 - \frac{E_1[5^4]}{4} \right) \right] \right\}.$$

Consequently, the first step must be to compute $E_1[5^4]$. Using the principles of 4.4

$$(6.3) \qquad E_1[5^4] \simeq E_1[E_1[5^2](.50), 5(.25), 5(.25)].$$

From Fig. 8, $E_1[5^2] = 7.55$. Therefore $E_1[5^4]$ is approximately equal to

$$E_1[7.55(.50), 5(.25), 5(.25)].$$

Now applying the same principle again,

$$(6.4) \qquad E_1[5^4] \simeq E_1[E_1[7.55(\tfrac{2}{3}), 5(\tfrac{1}{3})](.75), 5(.25)].$$

By the use of figures 7, 8, 9 and 10, graphical interpolation may be applied to find that $E_1[7.55(\tfrac{2}{3}), 5(\tfrac{1}{3})]$ is equal to 9.84. The approximation procedure now says that

$$(6.5) \qquad E_1[5^4] \simeq E_1[9.84(.75), 5(.25)].$$

Again applying the curves and using graphical interpolation for p_1 and p_2, $E_1[5^4] \simeq 11.88$.

Substituting this value in (6.2),

$$(6.6) \qquad E_4[5^5] \simeq \tfrac{5}{4} \{11.88 + E_3[2.71, 2.71, 2.71, 2.03]\}.$$

Now formula (4.35) must be applied to $E_3[2.71, 2.71, 2.71, 2.03]$, i.e.

$$E_3[2.71, 2.71, 2.71, 2 \; 3] \simeq$$

$$(6.7) \quad \tfrac{4}{3} \left\{ E_1[(2.71)^3] + E_2 \left[\left(\tfrac{3}{2} 2.71 - \frac{E_1[(2.71)^3]}{2} \right)^2, \quad \left(2.03 - \frac{E_1[(2.71)^3]}{3} \right) \right] \right\}.$$

$E_1[(2.71)^3]$ can be evaluated by the same method used for $E_1[5^4]$. This leads to the result

$$(6.8) \qquad E_3[2.71. 2.71, 2.71, 2.03] \simeq \tfrac{4}{3} \{4.40 + E_2[1.86, 1.86, .56]\}.$$

Once more applying (4.35)

$$E_2[1.86, 1.86, .56] \simeq$$

$$(6.9) \quad \tfrac{3}{2} \left\{ E_1[(1.86)^2] + E_1[(2 \cdot 1.86 - E_1[(1.86)^2]), \quad \left(.56 - \frac{E_1[(1.86)^2]}{2} \right) \right] \right\}.$$

$E_1[1.86, 1.86]$ is equal, by the curves, to 2.25.   Therefore

(6.10) $E_2[1.86, 1.86, .56] \simeq \frac{2}{3} \{2.25 + E_1[1.47, - .56]\}.$

However, since the convention is observed that a negative quantity is replaced by zero,

(6.11) $E_1[1.47, - .56] = E_1[1.47, 0] = 0.$

Now working back through these various expressions,

(6.12) $E_4[5^5] \simeq \frac{5}{4} [11.88 + \frac{4}{3} [4.40 + \frac{3}{2} [2.25 + 0]]] = 27.81.$

From Table 2 it can be seen that the percentage errors for this approximation to $E_4[5^5]$, corresponding to the 95% confidence limits for this quantity, are -4.0% and -9.9%.

This example has illustrated most of the situations which will arise in the use of the approximations of this paper.

6.4. *Miscellaneous approximation formulae useful for computation.* There exists a relatively simple approximation to $E_1[k_1(p_1), k_2(p_2)]$, $p_1 + p_2 = 1$, when p_2 is near one.   Using (3.8) and making some obvious simplifications, one obtains

$$E_1[k_1(p_1), k_2(p_2)] = \frac{k_2}{p_2} + \frac{1}{p_2} \cdot \frac{(k_1 + k_2)!}{(k_1 - 1)!(k_2 - 1)!} \frac{1}{p_1} \int_0^{p_1} t^{k_1-1}(1 - t)^{k_2-1}(t - p_1) \, dt.$$

Since p_1 is near zero, $(1 - t)$ can be replaced by one, and the result is obtained that

$$E_1[k_1(p_1), k_2(p_2)] \simeq \frac{k_2}{p_2} - \frac{1}{p_2} p_1^{k_1} \frac{(k_1 + k_2)!}{(k_1 + 1)!(k_2 - 1)!}.$$

An approximation to the Incomplete Beta-Function, given by Tukey and Scheffé [8], may also prove useful at times.   The expression, changed slightly by those authors since publication, is

$$I_b(n - r + 1, r) \simeq 1 - \frac{1}{2\Gamma(r)} \int_0^{\chi_a^2} \left(\frac{\chi^2}{2}\right)^{r-1} e^{-(\chi^2/2)} \, d\chi^2,$$

where

$$\chi_a^2 = 2r \left[\frac{(1 - b)\dfrac{n + \frac{1}{2}}{r} - 1}{\sqrt{b}}\right] + 2r.$$

The right hand side of the first expression will be recognized as the χ^2 distribution with $2r$ degrees of freedom.   In the event that the tables of χ^2 are not adequate for the application of these expressions, the approximation of Wilson and Hilferty [10] should be used.   This approximation states that $(\chi^2/\nu)^{\frac{1}{3}}$ where ν is the number of degrees of freedom, is approximately normally distributed with mean $1 - 2/(9\nu)$ and variance $2/(9\nu)$, for large ν.

7. Acknowledgements. The author wishes to express his grateful appreciation for the many helpful comments and suggestions received from Professors W. G. Cochran, A. M. Mood, J. W. Tukey and S. S. Wilks.

REFERENCES

[1] R. A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1943, Table XXII.

[2] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems with applications", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[3] J. B. S. HALDANE, "On a method of estimating frequencies", *Biometrika*, Vol. 33 (1945), pp. 222–225.

[4] P. S. LAPLACE, *Théorie Analytique des Probabilités*, Mme. Vᵉ Courcier, Paris, 1820, pp. 194–219.

[5] P. J. McCARTHY, *Approximate Solutions for Means and Variances in a Certain Class of Box Problems*, unpublished thesis, Library, Princeton University, 1946.

[6] KARL PEARSON, *Tables of the Incomplete Beta-Function*, The "Biometrika" Office, London, 1934.

[7] L. H. C. TIPPETT, *Random Sampling Numbers*, Cambridge Univ. Press, 1927.

[8] J. W. TUKEY, AND H. SCHEFFÉ, "A formula for sample sizes for population tolerance limits", *Annals of Math. Stat.*, Vol. 15 (1944), p. 217.

[9] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill, 1937, p. 181.

[10] E. B. WILSON AND M. M. HILFERTY, "The distribution of chi-square", *Proc. Nat. Acad. Sci.*, Vol. 17 (1931), pp. 684–688.

THE DISTRIBUTION OF THE RANGE[1]

By E. J. Gumbel

Brooklyn College, N. Y.

1. Summary. The asymptotic distribution of the range w for a large sample taken from an initial unlimited distribution possessing all moments is obtained by the convolution of the asymptotic distribution of the two extremes. Let α and u be the parameters of the distribution of the extremes for a symmetrical variate, and let $R = \alpha(w - 2u)$ be the reduced range. Then its asymptotic probability $\Psi(R)$ and its asymptotic distribution $\psi(R)$ may be expressed by the Hankel function of order one and zero. A table is given in the text.

The asymptotic distribution $g(w)$ of the range proper is obtained from $\psi(R)$ by the usual linear transformation. The initial distribution and the sample size influence the position and the shape of the distribution of the range in the same way as they influence the distribution of the largest value. If we take the parameters from the calculated means and standard deviations, the asymptotic distribution of the range gives a good fit to the calculated distributions for normal samples from size 6 onward. Consequently the distribution of the range for normal samples of any size larger than 6 may be obtained from the asymptotic distribution of the reduced range.

The asymptotic probabilities and the asymptotic distributions of the mth range and of the range for asymmetrical distributions are obtained by the same method and lead to integrals which may be evaluated by numerical methods.

2. Introduction. For any initial distribution, and any sample size n, the distribution of the range may easily be written down in the form of an integral. However, for many given initial distributions the integration can be carried out—if at all—only for very small sample sizes, say $n = 2$ or $n = 3$. For larger samples, complicated numerical calculations have to be made, and there is no way of obtaining the distribution for $n + 1$ observations from the distribution for n observations.

Our object is to obtain the *asymptotic* distribution of the range. Nothing is supposed to be known about the initial distribution, except that it is of the exponential type [9] which assures that it is unlimited in both directions, and possesses all moments. It will be shown that this condition is sufficient for the existence of an asymptotic distribution of the range.

With increasing samples sizes the distribution of the range may approach its asymptotic form in a quick, or in a slow way. This behavior depends upon the nature of the initial distribution. Two examples for this approach will be shown.

---

3. The exact distribution of the range. Let $\varphi(x)$ be any initial distribution, $\Phi(x)$ the probability of a value equal to, or less than, x. Then, for samples of size n, the joint distribution $\mathfrak{w}_n(x_1, x_n)$ of the smallest value x_1 and the largest value x_n is

(1) $$\mathfrak{w}_n(x_1, x_n) = n(n-1)\varphi(x_1)(\Phi(x_n) - \Phi(x_1))^{n-2}\varphi(x_n).$$

The distribution $g_n(w_n)$ of the range w_n defined by

(2) $$x_n = x_1 + w_n$$

is obtained by integrating over all values $x_1 \leq x_n$ whence

(3) $$g_n(w_n) = n(n-1) \int_{-\infty}^{+\infty} (\Phi(x + w_n) - \Phi(x))^{n-2} \varphi(x + w_n)\varphi(x) \, dx,$$

where the index 1 has been dropped. The probability $G_n(w_n)$ for the range to be equal to, or less than, w_n is obtained by integration of (3), whence, by reversing the order of integration,

$$G_n(w_n) = n \int_{-\infty}^{+\infty} \int_0^{w_n} (n-1)(\Phi(x + w_n) - \Phi(x))^{n-2} \, d\Phi(x + w_n) \, d\Phi(x),$$

or, after integration,

$$G_n(w_n) = n \int_0^1 (\Phi(x + w_n) - \Phi(x))^{n-1} \, d\Phi,$$

a formula to which Prof. H. Hotelling has drawn my attention. The beauty of this formula is completely marred by the facts that, in general, we cannot express $\Phi(x + w_n)$ by $\Phi(x)$, and that the numerical integration is lengthy and tiresome.

The problem of the range for the normal distribution was first raised twenty five years ago by L. von Bortkiewicz [1, 2]. For $n = 2$ and $n = 3$ the distribution of the normal range may be written down explicitly [12, 13]. For larger normal samples up to $n = 20$, E. S. Pearson [16] and H. O. Hartley [10] have calculated numerical tables of the probability of the range. L. H. C. Tippett [20] has calculated the mean, the standard deviation, and the moment quotients for the range of the normal distribution up to $n = 1000$. He gave formulae for the moments in the form of integrals. Finally "Student" [18] reproduced the distribution of the range for small samples, $n = 2, 3, 4, 5, 6, 10$, by Pearson's type I, and gave a formula for large samples $n = 20, 60$, based on Pearson's type VI, a procedure which is purely empirical and, therefore, unsatisfactory for theoretical purposes. A good resumé of the present knowledge about the range is given in Karl Pearson's Tables [17].

All these studies are confined to the normal distribution and allow no conclusion about the asymptotic distribution of the range. According to Kendall [11] it is not known whether such forms exist and what they are. This question may at once be answered for a special case. If the distribution is limited to the left (or to the right), the asymptotic distribution of the range is equal to the asymp-

totic distribution of the largest (smallest) value. The asymptotic distribution of the range exists provided that an asymptotic distribution of the largest (smallest) value exists. For the exponential distribution, and for initial distributions of the Pareto type, for example, the asymptotic distribution of the range is equal to the asymptotic distribution of the largest value. The asymptotic distribution of the range for the rectangular distribution has been derived by A. G. Carlton [3].

4. The asymptotic distribution of the reduced range for a symmetrical variate. Instead of the procedures mentioned in the last paragraph, let us consider a large sample. It is generally assumed that the smallest and the largest values are independent in that case. L. H. C. Tippett [20] has shown that the correlation between the extremes is negligible for the normal distribution and for sample sizes $n \geqq 200$. In a previous note [9] it has been shown that independence holds for large samples and for initial distributions of the exponential type unlimited in both directions and possessing all moments. Then the joint distribution (1) splits into the product of the asymptotic distribution $f_1(x_1)$ of the smallest value x_1 and the asymptotic distribution $f_n(x_n)$ of the largest value x_n

$$(4) \qquad\qquad \mathfrak{w}(x_1, x_n) = f_1(x_1) \cdot f_n(x_n).$$

If, furthermore, the initial distribution is symmetrical about zero, the two asymptotic distributions are

$$(5) \quad f_1(x_1) = \alpha \exp[\alpha(x_1 + u) - e^{\alpha(x_1 + u)}]; \; f_n(x_n) = \alpha \exp[-\alpha(x_n - u) - e^{-\alpha(x_n - u)}].$$

These asymptotic distributions and the corresponding probabilities are traced, in a reduced scale, on Graphs (1) and (2).

Since the two parameters u and α will exist also in the asymptotic distribution of the range, their nature must briefly be explained. The value u is defined as the solution of

$$(6) \qquad\qquad \Phi(u) = 1 - \frac{1}{n}.$$

Since

$$(6') \qquad\qquad n(1 - \Phi(u)) = 1,$$

the largest value u may be called the *expected* largest value. It differs, of course, from the mean of the largest value. It has been shown [6] that u increases as a function of the logarithm of n, the function depending upon the initial distribution.

Criteria for the approach of the distribution of the largest value toward its asymptotic form have been given by R. A. Fisher and L. H. C. Tippett [4].

For our purpose it is sufficient to consider whether n is so large that u is very near to the most probable largest value $\tilde{x}_n$ obtained from

$$(7) \qquad \frac{n-1}{\Phi(\tilde{x}_n)} \, \varphi(\tilde{x}_n) = -\frac{\varphi'(\tilde{x}_n)}{\varphi(\tilde{x}_n)}.$$

If

$$\tilde{x}_n \approx u$$

holds with sufficient approximation, $2u$ may be interpreted as the range of the modes for an initial symmetrical distribution.

The parameter α defined by

$$(8) \qquad \alpha = \frac{\varphi(u)}{1 - \Phi(u)}$$

also is a function of n. Three cases have to be distinguished: In the first case, α is a constant, or converges with n toward a constant different from zero. In the second (and third) case, α increases with n without limit (decreases with n toward zero). The three cases correspond to three classes of initial distributions of the exponential type. The function α is related to the asymptotic standard error of the largest, and of the smallest value by

$$(9) \qquad \alpha^2 \sigma_n^2 = \alpha^2 \sigma_1^2 = \frac{\pi^2}{6}.$$

If α increases (decreases) with n, or is independent of n, the standard error of the largest value decreases (increases) with the sample size, or is independent of it. This behavior has nothing to do with the fact that the standard error of the mean decreases, of course, with an increasing number of samples.

The determination of the constants u and α from equations (6), (7), (8) is based on the knowledge of the initial distribution and the sample size n from which we take the largest observation. This method cannot be used in many practical applications: 1) It may happen that the initial distribution, or the parameters it contains, are unknown. Therefore the parameters of the largest value cannot be obtained from it. 2) The initial distribution might be known, but the number of observations is insufficient to warrant this procedure, because the most probable largest value $\tilde{x}_n$ differs from the expected value u. In these cases the parameters u and α have to be estimated from the observed distribution of the largest value alone. A similar procedure will be used for the range in paragraph 7.

From (4) and (5) the joint asymptotic distribution $w(x, w)$ of the smallest value x_1 and the range w becomes

$$w(x_1, w) = \alpha^2 \exp[-\alpha(w - 2u) - e^{\alpha(x_1 + u)} - e^{-\alpha(x_1 + w - u)}].$$

The asymptotic distribution $g(w)$ of the range alone is, dropping the index 1,

$$(4') \qquad g(w) = \alpha^2 e^{-\alpha(w - 2u)} \int_{-\infty}^{+\infty} \exp[-e^{\alpha(x + u)} - e^{-\alpha(x + w - u)}] \, dx.$$

This distribution contains the two parameters α and u existing in the asymptotic distribution of the largest value. To eliminate the two parameters, a reduced range R is introduced by

$$(10) \qquad\qquad R = \alpha(w - 2u).$$

The range w is a positive variate unlimited toward the right. The reduced range R is also unlimited toward the right yet limited toward the left by

$$(10') \qquad\qquad R \geq -2\alpha u.$$

The reduced range is not related to one of the averages of the range. It is the range minus the range of the modes divided by a factor which is proportional to the standard error of the extreme value. The distribution $\psi(R)$ of the reduced range R, and the distribution $g(w)$ of the range w are related by

$$(11) \qquad\qquad \psi(R) = \frac{1}{\alpha} g(w),$$

subject to restriction (10'), whereas the probability $\Psi(R)$ of the reduced range to be equal to, or less than R is equal to the corresponding expression $G(w)$ for the range proper

$$(11') \qquad\qquad \Psi(R) = G(w).$$

For the integration in (4') we put

$$\alpha(x + w - u) = -y$$

whence, from (10),

$$\alpha(x + u) = -y - R.$$

The asymptotic distribution of the reduced range becomes

$$(12) \qquad\qquad \psi(R) = e^{-R} \int_{-\infty}^{+\infty} \exp[-e^y - e^{-y-R}] \, dy$$

and the asymptotic probability $\Psi(R)$ of the range is

$$(13) \qquad\qquad \Psi(R) = \int_{-\infty}^{+\infty} \exp[y - e^y - e^{-y-R}] \, dy$$

an expression which may easily be verified by differentiation.

The asymptotic formulas (12) and (13) hold for any initial symmetrical distribution of the exponential type, for example, for the normal and the logistic distribution (see par. 7). The mean reduced range $\bar{R}$ and the higher moments of the reduced range are easily obtained from the mean $\bar{w}$, the variance σ_w^2, and

the invariants λ_ν of order ν of the range proper w given in a previous paper [8]. They are

(14)
$$\bar{w} = 2u + \frac{2\gamma}{\alpha}\;;\qquad \sigma_w^2 = \frac{\pi^2}{3\alpha^2}$$

(15)
$$\lambda_\nu = \frac{2(\nu-1)!}{\alpha^\nu}\sum_{k=1}^{\infty}\frac{1}{k^\nu}\;;\qquad \nu \geqq 2$$

where γ stands for Euler's constant.

Consequently the mean $\bar{R}$, the variance σ_R^2 and the invariants λ_ν of the reduced range are

(16)
$$\bar{R} = 2\gamma;\quad \sigma_R^2 = \frac{\pi^2}{3}\;;\quad \lambda_\nu = 2(\nu-1)!\sum_{k=1}^{\infty}\frac{1}{k^\nu}\;;\qquad \nu \geqq 2$$

Equation (14) leads to an interpretation of the reduction (10) which may be written

$$R = \alpha(w - \bar{w}) + 2\gamma$$

or

(14′)
$$R = \frac{\pi}{\sqrt{3}}\frac{w - \bar{w}}{\sigma_w} + 2\gamma$$

Thus the transformation (10) is a linear function of the standard transformation $(w - \bar{w})/\sigma_w$ usual in statistics.

5. The probability of the range as a Bessel function. The integrals (12) and (13) may be evaluated by numerical procedures, since tables of the function $\exp(-e^{-\nu})$ are easily calculated. However, it turned out to be simpler to relate these integrals to the solution of a differential equation. The derivative $\psi'(R)$ of the distribution (12) is

$$\psi'(R) = -\psi(R) + e^{-R}\int_{-\infty}^{+\infty}\exp[-y - R - e^y - e^{-y-R}]\,dy$$

The integral is equal to the probability $\Psi(R)$ since the transformation

$$y + R = -z$$

leads to

$$\int_{-\infty}^{+\infty}\exp[-y - R - e^y - e^{-y-R}]\,dy = \int_{-\infty}^{+\infty}\exp[z - e^{-z-R} - e^z]\,dz$$

Consequently the probability $\Psi(R)$ is subject to the differential equation

(17)
$$\Psi'' + \Psi' - e^{-R}\Psi = 0.$$

The mode of the reduced range is a fixed value $\tilde{R}$ such that

(18) $$\psi(\tilde{R}) = e^{-\tilde{R}}\psi(\tilde{R}).$$

Mr. W. Wasow (Swarthmore College) has drawn my attention to the fact that the probability $\Psi(R)$ of the range can be expressed in terms of a Bessel function.[2] To obtain this simplification of the differential equation we introduce a new positive variable z by

(19) $$z = 2e^{-R/2}$$

and a new function U by

(20) $$\Psi = U \cdot z.$$

The boundary conditions are

(21) $$z = 0, \Psi = 1; \quad z = \infty; \quad \Psi = 0.$$

The first derivative becomes, from (19)

$$\frac{d\Psi}{dR} = -\frac{z}{2}\frac{d\Psi}{dz}$$

whence, from (20)

$$\frac{d\Psi}{dR} = -\frac{z}{2}\left(U + z\frac{dU}{dz}\right).$$

The second derivative becomes, by the same procedure

$$\frac{d^2\Psi}{dR^2} = -\frac{z}{2}\frac{d}{dz}\left(-\frac{zU}{2} - \frac{z^2}{2}\frac{dU}{dz}\right).$$

The second member may be written

$$\frac{z}{2}\left(\frac{U}{2} + \frac{3z}{2}\frac{dU}{dz} + \frac{z^2}{2}\frac{d^2U}{dz^2}\right) = \frac{zU}{4} + \frac{3z^2}{4}\frac{dU}{dz} + \frac{z^3}{4}\frac{d^2U}{dz^2}.$$

Thus the differential equation (17) is now

$$\frac{z^3}{4}\frac{d^2U}{dz^2} + \frac{3z^2}{4}\frac{dU}{dz} - \frac{z^2}{2}\frac{dU}{dz} + \frac{zU}{4} - \frac{zU}{2} - \frac{z^3}{4}U = 0.$$

Multiplication by $4z^{-1}$ leads to

(21') $$z^2\frac{d^2U}{dz^2} + z\frac{dU}{dz} - (z^2 + 1)U = 0.$$

This is one of the classical Bessel differential equations of order 1. In the notation used by the British Tables [14] (pp. 264 and 213) one of the solutions is

(22) $$U(z) = K_1(z),$$

---

where $K_1(z)$, the modified Bessel function of the second kind (Hankelfunction) is defined by

$$(23) \quad K_1(z) = (\gamma - \lg 2 + \lg z) \sum_0^\infty \frac{1}{\nu!(\nu+1)!} \left(\frac{z}{2}\right)^{2\nu+1}$$

$$+ \frac{1}{z} - \sum_1^\infty \frac{1}{(\nu-1)!\,\nu!} \left(\frac{z}{2}\right)^{2\nu-1} \left(1 + \frac{1}{2} + \cdots + \frac{1}{\nu} - \frac{1}{2\nu}\right).$$

The relation between the functions $K_\nu(z)$ and the Hankelfunction $H_\nu^{(1)}(z)$ is

$$(23a) \qquad\qquad K_\nu(z) = \frac{\pi}{2} i^{\nu+1} H_\nu^{(1)}(iz).$$

The asymptotic probability for the range is, from (20) and (22),

$$(24) \qquad\qquad \Psi(R) = zK_1(z)$$

or, from (19)

$$(25) \qquad\qquad \Psi(R) = 2e^{-R/2} K_1(2e^{-R/2}).$$

This is the only Bessel function satisfying the boundary conditions (21). The asymptotic probability $\Psi(R)$ of the range may be written finally from (25), (23) and (10)

$$(25a) \qquad 1 - \Psi(R) = \sum_1^\infty \frac{\exp(-R\nu)}{\nu!\,(\nu-1)!} \left(R - 2\gamma + 2S_\nu - \frac{1}{\nu}\right)$$

where

$$S_0 = 0; \quad S_\nu = \sum_{\lambda=1}^\nu \frac{1}{\lambda}$$

The distribution

$$\psi(R) = \frac{d\Psi(R)}{dz} \cdot \frac{dz}{dR}$$

of the reduced range R is, from (24) and (19)

$$\psi(R) = -\frac{z}{2}\,(K_1(z) + zK_1'(z)).$$

Now, the derivative $K_1'(z)$ is linked to the modified Bessel function $K_0(z)$ of the second kind and of order zero by

$$zK_1'(z) = -K_1(z) - zK_0(z).$$

Consequently the distribution is

$$(26) \qquad\qquad \psi(R) = \frac{z^2}{2} K_0(z)$$

or, from (19),

$$(27) \qquad \psi(R) = 2e^{-R}K_0(2e^{-R/2})$$

where the function $K_0(z)$ is defined by

$$(28) \qquad K_0(z) = -(\gamma - \lg 2 + \lg z) \sum_0^\infty \left(\frac{z}{2}\right)^{2\nu} \frac{1}{\nu!\,\nu!}$$

$$+ \sum_1^\infty \left(\frac{z}{2}\right)^{2\nu} \frac{1}{\nu!\,\nu!} \left(1 + \frac{1}{2} + \cdots \frac{1}{\nu}\right).$$

Finally the asymptotic distribution $\psi(R)$ of the reduced range may be written from (27) and (28)

$$(28a) \qquad \psi(R) = \sum_0^\infty \frac{\exp\left[-(\nu+1)R\right]}{\nu!\,\nu!} (R - 2\gamma + 2S_\nu)$$

We first investigate the analytic behavior and the order of magnitude of the probability $\Psi(R)$ and the distribution $\psi(R)$ for large negative, and large positive values of the reduced range, i.e. for large and small values of the positive variable z. If z is so large that

$$(29) \qquad z^{-3} = \frac{e^{(3R/2)}}{8} \ll 1$$

the expressions for $K_1(z)$ and $K_0(z)$ become [14], p. 271,

$$K_1(z) = \sqrt{\frac{\pi}{2z}}\, e^{-z} \left(1 + \frac{3}{8z} - \frac{15}{128z^2}\right)$$

$$K_0(z) = \sqrt{\frac{\pi}{2z}}\, e^{-z} \left(1 - \frac{1}{8z} + \frac{9}{128z^2}\right)$$

The probability $\Psi(R)$ becomes, from (24) and (19),

$$(25') \qquad \Psi(R) = \sqrt{\pi}\, \exp\left[-\frac{R}{4} - 2e^{-(R/2)}\right]\left(1 + \frac{3}{16}\, e^{R/2} - \frac{15}{512}\, e^R\right).$$

The condition (29) holds, say, for $R = -4$. The numerical calculation leads, for $\Psi(-4)$, to the order of magnitude 10^{-6}.

In the same way, the distribution $\psi(R)$ becomes, from (26) and (19), for large negative reduced ranges

$$(27') \qquad \psi(R) = \sqrt{\pi}\, \exp\left[-\frac{3R}{4} - 2e^{-(R/2)}\right]\left(1 - \frac{e^{R/2}}{16} + \frac{9}{512}\, e^R\right).$$

This expression cannot be obtained from (25') since the approximations for $K_0(z)$ and $K_1(z)$ used do not fulfill the relations between the derivatives given above. The order of magnitude of $\psi(-4)$ is 10^{-5}.

Thus the probability $\Psi(R)$ and the distribution $\psi(R)$ may be neglected for $R \leqq -4$. This removes the importance of the lower limit $R \geqq -2\alpha u$ stated

in (10'). If $\alpha u \geqq 2$, the distribution of the range may be dealt with as if it were practically unlimited toward the left.

For large positive reduced ranges to which correspond small values of z, say

(29') $$z^3 = 8e^{-(3R/2)} \ll 1$$

the Bessel functions $K_1(z)$ and $K_0(z)$ become, from (23) and (28)

(23') $$K_1(z) = \left(\gamma + \lg \frac{z}{2}\right)\left(\frac{z}{2} + \frac{z^3}{16}\right) + \frac{1}{z} - \left(\frac{z}{4} + \frac{5z^3}{64}\right)$$

(28') $$K_0(z) = -\left(\gamma + \lg \frac{z}{2}\right)\left(1 + \frac{z^2}{4}\right) + \frac{z^2}{4} + \frac{3z^4}{128} .$$

In this case we are interested to know how far the probability $\Psi(R)$ differs from unity. Consequently we calculate $1 - \Psi(R)$ and obtain, from (24) and (23')

$$1 - \Psi(R) = -\frac{z^2}{2}\left[\left(\gamma + \lg \frac{z}{2}\right)\left(1 + \frac{z^2}{8}\right) - \frac{1}{2} - \frac{5z^2}{32}\right].$$

The right side becomes, from (19)

$$2e^{-R}\left[\left(-\gamma + \frac{R}{2}\right)\left(1 + \frac{e^{-R}}{2}\right) + \frac{1}{2} + \frac{5}{8}e^{-R}\right]$$

$$= e^{-R}\left[(R - 2\gamma)\left(1 + \frac{e^{-R}}{2}\right) + 1 + \frac{5}{4}e^{-R}\right]$$

or

$$e^{-R}\left[R - 2\gamma + 1 + \frac{e^{-R}}{2}\left(R - 2\gamma + \frac{5}{2}\right)\right] = e^{-R}(R - 2\gamma + 1)\left(1 + \frac{e^{-R}}{2}\right) + \frac{3e^{-2R}}{4} .$$

If R is so large that

$$e^{-R} \ll 1$$

we simply have

(25'') $$1 - \Psi(R) = e^{-R}(R - 2\gamma + 1) .$$

For example, for $R = 10$, the preceding condition is satisfied and $1 - \Psi(R)$ is of the order 5.10^{-4}.

In the same manner we calculate the density of probability $\psi(R)$ for large reduced ranges. From (26), (19) and (28') we obtain

$$\psi(R) = 2e^{-R}\left[\left(\frac{R}{2} - \gamma\right)(1 + e^{-R}) + e^{-R} + \frac{3}{8}e^{-2R}\right].$$

By neglecting $e^{-2R} \ll R$, the right side becomes

$$e^{-R}[(R - 2\gamma)(1 + e^{-R}) + 2e^{-R}] = e^{-R}[R - 2\gamma + e^{-R}(R - 2\gamma + 2)]$$

whence

$$\psi(R) = e^{-R}(R - 2\gamma)(1 + e^{-R}) + 2e^{-2R}.$$

In first approximation we obtain

(27″) $\psi(R) = e^{-R}(R - 2\gamma)$

a formula which may also be derived directly from (25″). The density of probability is of the order 10^{-4} for $R = 10$.

From the formulae (25′) and (27′) valid for large negative values of R, and from the formulae (25″) and (27″) valid for large positive values of R follow the boundary conditions

$$\lim_{R \to -\infty} \frac{\psi(R)}{\Psi(R)} = e^{-(R/2)} \; ; \; \lim_{R \to +\infty} \frac{\psi(R)}{1 - \Psi(R)} = \frac{R - 2\gamma}{R - 2\gamma + 1}$$

For the construction of tables of the distribution $\psi(R)$ and the probability $\Psi(R)$ of the reduced range it is sufficient to consider the interval

$$-3 < R < 10.$$

The two functions $K_1(z)$ and $K_0(z)$ have been tabulated [14] and [19].   Hence the probability and the distribution could be calculated from such tables of the Bessel functions. This procedure, however, was only used to obtain boundary values.   The tables I and Ia are based on computations made in the Calculation and Ballistics Department at the Naval Proving Ground Dahlgren by stepwise integration of the differential equation (17) using the special Relay Calculator of the International Business Machines Corporation.[3]

Table I gives the probability $\Psi(R)$ (col. 2) and the distribution $\psi(R)$ (col. 4) for the reduced ranges $-3 \leqq R \leqq 10.5$ in intervals $\Delta R = 0.5$.   The differences $\Delta\Psi$ given in col. 3 are taken from the original figures.

For different uses it is necessary to know the reduced range as a function of its probability.   This relation is shown in Table Ia.   The first column gives the probability, the first line gives the last decimal of this probability, and the cells give the reduced range corresponding to the probability obtained from the combination of the first column and the first line.   For example: The reduced range $R = -3.20$ corresponds to the probability $\Psi(R) = 0.0002$, and the reduced range $R = 10.44$ corresponds to the probability $\Psi(R) = 0.9997$.

This table may be used for obtaining the percentage points of the reduced range.   The mode $\hat{R}$, the median $\check{R}$ calculated by the Naval Proving Ground and the mean $\bar{R}$ obtained from (14) and (10) are

(30) $\hat{R} = 0.506366440;$ $\check{R} = 0.928597642;$ $\bar{R} = 1.154431330.$

A probability paper for the range may be constructed in the following way: The observed ranges w are plotted on the vertical axis; the reduced ranges R on a horizontal axis.   The abscissa shows the probabilities

$$\Psi(R) = G(w)$$

---

[3] The author wishes to express his sincere appreciation for the permission to use these computations.   The original tables give the probability and the distribution to 8 significant decimal places at intervals $\Delta R = 1/100$.   Lack of space prevents the reproduction of these tables.

TABLE I

Asymptotic Probability and Asymptotic Distribution of the Reduced Range

1	2	3	4
Reduced Range R	Probability $\Psi (R)$	Difference $\Delta\Psi$	Distribution $\psi (R)$
−3.0	.00050		.00212
		.00274	
−2.5	.00324		.01057
		.01032	
−2.0	.01356		.03386
		.02693	
−1.5	.04048		.07705
		.05251	
−1.0	.09299		.13419
		.08141	
− .5	.17440		.18969
		.10533	
.0	.27973		.22779
		.11821	
.5	.39794		.24075
		.11859	
1.0	.51654		.23021
		.10891	
1.5	.62545		.20346
		.09327	
2.0	.71872		.16898
		.07557	
2.5	.79429		.13360
		.05860	
3.0	.85289		.10157
		.04386	
3.5	.89675		.07483
		.03192	
4.0	.92867		.05375
		.02270	
4.5	.95136		.03783
		.01584	
5.0	.96721		.02618
		.01089	
5.5	.97810		.01787
		.00739	
6.0	.98549		.01205
		.00496	

TABLE I—*Concluded*

1	2	3	4
Reduced Range R	Probability Ψ (R)	Difference $\Delta\Psi$	Distribution ψ (R)
6.5	.99045		.00805
		.00330	
7.0	.99375		.00534
		.00218	
7.5	.99594		.00351
		.00143	
8.0	.99737		.00230
		.00093	
8.5	.99830		.00150
		.00061	
9.0	.99891		.00097
		.00039	
9.5	.99930		.00062
		.00025	
10.0	.99955		.00040
		.00016	
10.5	.99972		.00026

corresponding to the reduced ranges R. If the observations follow the theory, the observed ranges are scattered around the straight line

$$(10') \qquad\qquad w = 2u + \frac{R}{\alpha}$$

If the samples are drawn simultaneously, and if there is a constant interval of time between the drawings, this interval may be used as unit of time for the construction of the return periods $T(R)$ and $_1T(R)$ of a range equal to, or larger than (smaller than) R where

$$T(R) = \frac{1}{1 - \Psi(R)} \; ; \; _1T(R) = \frac{1}{\Psi(R)} \; .$$

The first (second) notion applies to the range above (below) the median. The return periods are shown in an upper parallel to the abscissa.

A scheme for this paper is given in Fig. 3. Such a paper will allow a graphical test for the fit of the observed ranges to our theory, and avoids any numerical calculations. Obviously this method may only be used if the initial distribution is symmetrical, unlimited, and of the exponential type, and if the sample size is so large that the asymptotic distribution holds.

6. The range, the midrange, and the extremes. The asymptotic distribution (27) of the reduced range was obtained by convolution of the asymptotic distributions (5) of the extremes. The same method leads to the asymptotic distribution of the reduced midrange [8]

$$(31) \qquad\qquad v = \alpha(x_1 + x_n).$$

TABLE IA

The Reduced Range R as Function of Its Probability $\Psi (R)$

Ψ (R)	0	1	2	3	4	5	6	7	8	9
.000	—	*	−3.20	−3.12	−3.05	−3.00	−2.96	−2.92	−2.89	−2.86
.00	—	−2.83	−2.64	−2.52	−2.43	−2.36	−2.30	−2.25	−2.20	−2.16
.0	—	−2.12	−1.84	−1.65	−1.51	−1.39	−1.28	−1.19	−1.10	−1.02
.1	−0.95	−0.88	−0.81	−0.75	−0.69	−0.63	−0.58	−0.52	−0.47	−0.42
.2	−0.37	−0.32	−0.27	−0.22	−0.18	−0.13	−0.09	−0.04	0.00	0.04
.3	0.09	0.13	0.17	0.22	0.26	0.30	0.34	0.38	0.43	0.47
.4	0.51	0.55	0.59	0.63	0.68	0.72	0.76	0.80	0.84	0.89
.5	0.93	0.97	1.02	1.06	1.10	1.15	1.19	1.24	1.28	1.33
.6	1.38	1.43	1.47	1.52	1.57	1.62	1.67	1.73	1.78	1.84
.7	1.89	1.95	2.01	2.07	2.13	2.19	2.26	2.33	2.40	2.47
.8	2.54	2.62	2.70	2.79	2.88	2.97	3.07	3.18	3.29	3.41
.9	3.54	3.69	3.85	4.03	4.23	4.46	4.75	5.11	5.61	6.45
.99	6.45	6.57	6.71	6.87	7.05	7.26	7.52	7.85	8.31	9.10
.999	9.10	9.22	9.35	9.50	9.67	9.88	10.12	10.44	*	*

* These values have not been calculated.

On the other hand, the asymptotic distributions of the reduced extremes are obtained by introducing the transformations

$$(32) \qquad\qquad y_1 = \alpha(x_1 + u); \qquad y_n = \alpha(x_n - u)$$

into formulas (5). It is interesting to compare these four distributions and four probabilities with each other. This is done in Figures 1 and 2. The probability and the distribution of the midrange are practically identical with the probability and distribution of the smallest value, for small values of the midrange, and become practically identical with the probability and distribution of the largest value for large values of the midrange. Fig. 2 shows that the asymptotic distribution of the reduced range is less asymmetrical than the asymptotic distributions of the reduced extremes.
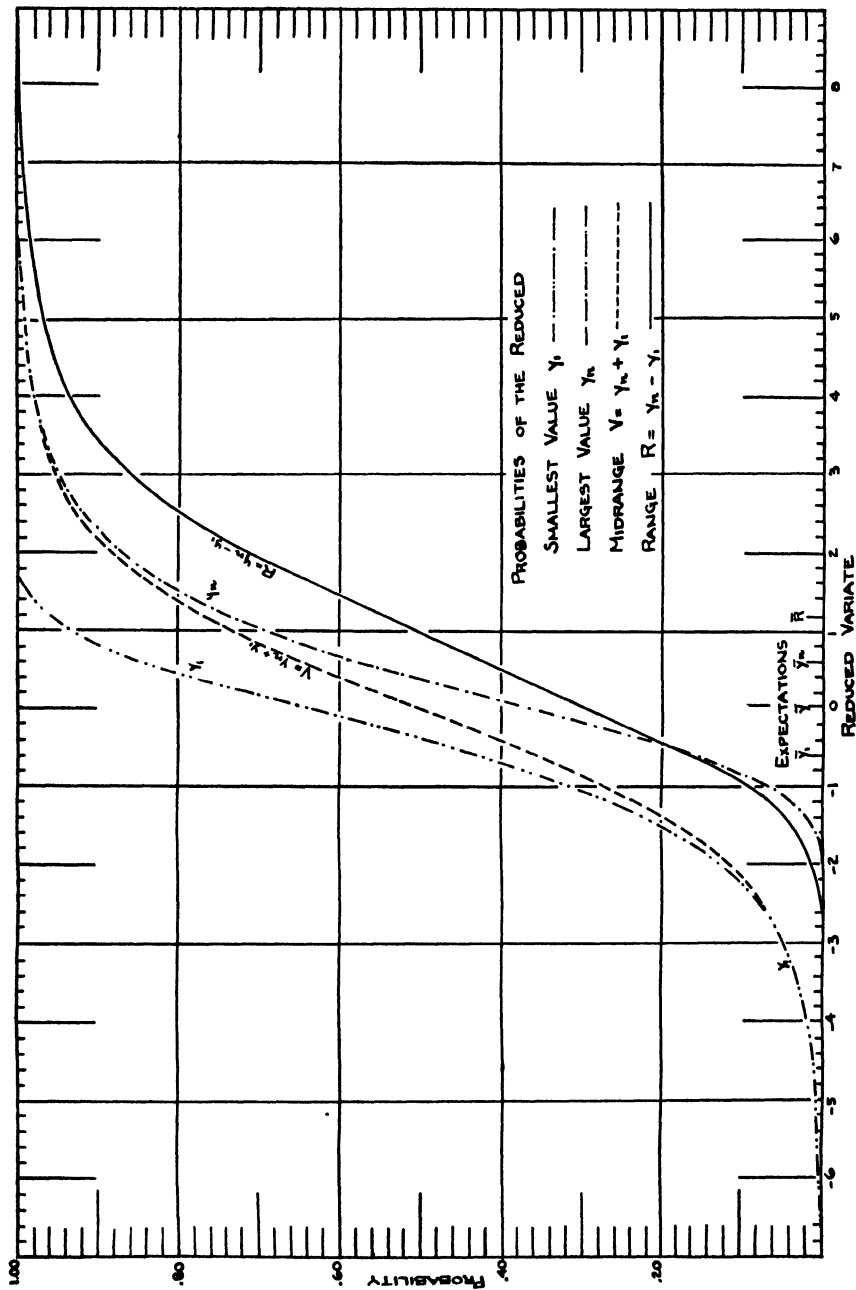
Fig. 1

Table II contains some characteristic values for these four asymptotic distributions. The first three columns are obtained from previous publications [6, 8]. The mean range is equal to the range of the means for the extremes. The median of the range is larger than the range from the median of the largest to the median of the smallest value. The mode of the range is slightly smaller than the mean of the largest value. These statements hold, of course, only for the reduced variates.
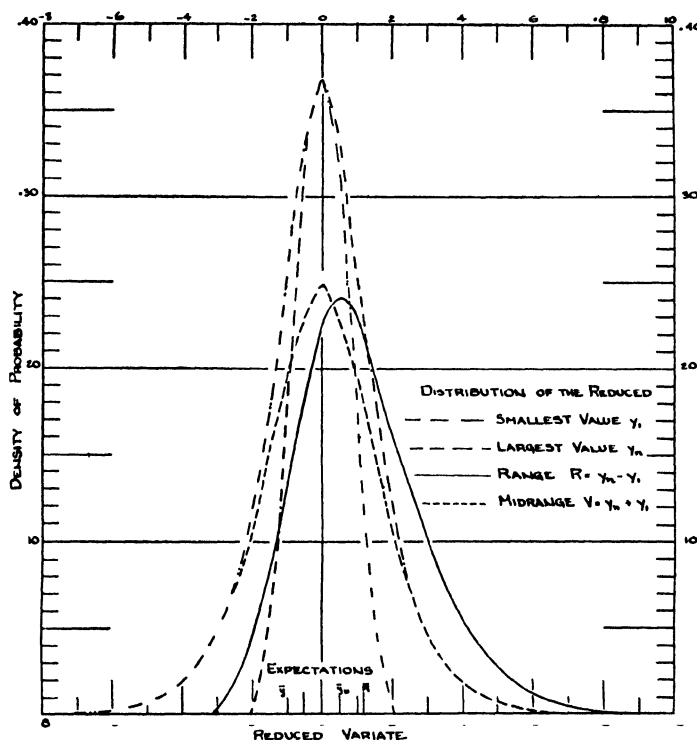
FIG 2

From the mode $\tilde{R}$ of the reduced range given in equation (30) and the transformation (10), the mode $\tilde{w}$ of the range itself is obtained as

$$\tilde{w} = 2u + \frac{\tilde{R}}{\alpha}$$

whereas the difference of the modes of the largest and of the smallest values is

$$\tilde{x}_n - \tilde{x}_1 = 2u.$$

Consequently

(33) $\tilde{w} = \tilde{x}_n - \tilde{x}_1 + \frac{\tilde{R}}{\alpha}$.

Scheme of a
Probability Paper for
the Range
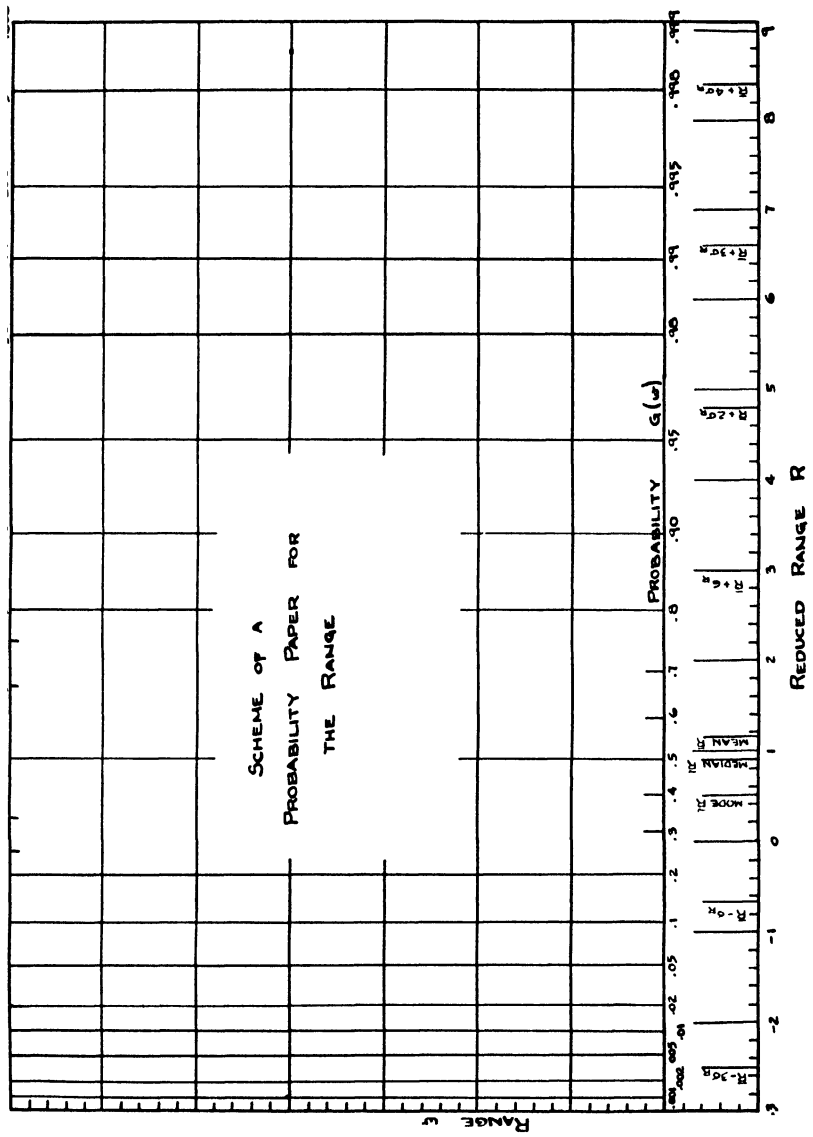
Fig. 3

For a symmetrical initial distribution of the exponential type the mode of the range converges toward the range of the modes of the smallest and of the largest value, provided that the parameter α increases without limit with the sample size. Thus this convergence does not hold for all symmetrical distributions.

The last two lines in Table II give the four probabilities corresponding to the intervals from the mean μ minus once (twice) the standard deviation σ, up to the mean plus once (twice) the standard deviation. The first probability for

TABLE II

Characteristics for the 4 Asymptotic Reduced Distributions

1 Characteristic	2 Largest Value	3 Smallest Value	4 Midrange	5 Range
Mode	0	0	0	.506
Expectation	$\gamma = .57722$	$= -.57722$	0	$2\gamma = 1.15444$
Median	$-\lg\lg 2 = .36651$	$= -.36651$	0	.929
Seminvariant char. function	$\Gamma(1 - t)$	$\Gamma(1 + t)$	$\Gamma(1 - t) \cdot \Gamma(1 + t)$	$\Gamma^2(1 - t)$
Variance	$\dfrac{\pi^2}{6} = 1.64493$	$= 1.64493$	$\dfrac{\pi^2}{3}$	$= 3.28986$
First + second mo- ment quotient	$\beta_1 = 1.29857$ $\beta_2 = 5.4$	-1.29857 5.4	0 4.2	$.64928$ 4.2
95% Probability	2.97	1.10	2.94	4.46
99% Probability	4.60	1.53	4.60	6.45
$F(\mu + \sigma) - F(\mu - \sigma)$	.72	.72	.72	.71
$F(\mu + 2\sigma) - F(\mu - 2\sigma)$	.90	.90	.95	.95

the four distributions is about the same as for the normal distribution. The second . probability for the range and the midrange is about the same as for the normal one.

7. The asymptotic distribution of the range for a symmetrical variate. The asymptotic distribution of the range R is, of course, independent of the sample size, and parameter-free. Both statements do not hold for the distribution $g(w)$ of the range proper which is, from (11)

$$(34) \qquad g(w) = \alpha\psi[\alpha(w - 2u)].$$

In this formula, the range is expressed in the same units as the initial variate. The parameters α and u are functions of the sample size n, the function depending

upon the initial distribution. From equations (6), (8), (14) follows that an increase of the sample size has two influences on the distribution of the range. The increase of the parameter u shifts the distribution toward the right without changing its form, whereas the parameter α influences the shape of the distribution. If α increases (decreases) with n, the distribution of the range shrinks (spreads) with increasing sample size. If α is independent of n, an increase of the sample size does not change the shape of the distribution. Only in the first case may we increase the precision of the range by increasing the sample size. The two parameters thus influence the range in the same way as they influence the extreme values.

To use equation (34) for a given initial distribution and a given sample size, we have to determine the expected largest value u and the parameter α as functions of n. We may use the definitions (6), (7), (8) if the initial distribution is known and of the exponential type, and if the sample size is so large that the most probable largest value is sufficiently near to the solution of (7).

As a first example, consider the so-called logistic distribution. This probability is

$$(35) \qquad\qquad \Phi(x) = (1 + e^{-x})^{-1}.$$

The initial distribution is

$$(35') \qquad\qquad \varphi(x) = \Phi(x)(1 - \Phi(x))$$

and the derivative is

$$(35'') \qquad\qquad \varphi'(x) = \Phi(x)(1 - \Phi(x))(1 - 2\Phi(x)).$$

Equation (6) becomes

$$1 + e^{-u} = \frac{n}{n - 1}$$

whence the expected largest value

$$(36) \qquad\qquad u = \lg(n - 1).$$

The most probable largest value $\tilde{x}_n$ for n observations is obtained from (7). This equation becomes, from equation (35)

$$(n - 1)(1 - \Phi(\tilde{x}_n)) = -1 + 2\Phi(\tilde{x}_n)$$

whence

$$\Phi(\tilde{x}_n) = \frac{n}{n + 1}$$

Equation (35) leads to the most probable largest value

$$(36') \qquad\qquad \tilde{x}_n = \lg n.$$

Even for n as small as 30 the difference between $\tilde{x}_n$ and u is less than 1%. Consequently the asymptotic form of the distribution of the range may be used even for small samples. The two parameters are

$$(37) \qquad\qquad u = \lg n; \qquad \alpha = \frac{n}{n + 1}.$$

Since α converges toward unity, an increase of the sample size shifts the distribution of the range toward the right without influencing its shape: the precision of any estimate made from the range cannot be increased by increasing the sample size.

The characteristic ranges introduced in paragraph 5 are obtained immediately: the mean $\bar{w}$, the mode $\tilde{w}$, the median range $\hat{w}$ and the ranges $w_{.95}$ and $w_{.99}$

$$\bar{w} = \lg n + 1.154; \quad \tilde{w} = \lg n + .506;$$

$$\hat{w} = \lg n + .929; \quad w_{.95} = \lg n + 4.46; \quad w_{.99} = \lg n + 6.45$$

are parallel straight lines if traced as functions of the sample size n on semi-logarithmic paper.

For the normal distribution we cannot expect such simple results. Here, u and α can only be calculated as numerical functions of n although limiting forms of these functions are known. The parameter α increases with n, and the standard error of the range decreases without limit although very slowly. The logistic distribution belongs to the first, the normal distribution to the second class of initial distributions of the exponential type.

The probabilities and the distributions of the range for normal samples of size 5, 10, and 20 as calculated by E. S. Pearson and H. O. Hartley [16] are traced in Figures 4 and 5. Our aim is to trace the corresponding asymptotic probabilities and distributions in order to see how far the asymptotic ranges differ from the exact ones. However, we have first to settle the preliminary question how far the most probable largest value $\tilde{x}_n$ differs from the expected largest value u. The most probable largest value $\tilde{x}_n$ is obtained from (7) which becomes, for the normal distribution,

$$(38) \qquad \tilde{x}_n \Phi(\tilde{x}_n) = (n-1)\varphi(\tilde{x}_n).$$

The results $\tilde{x}_n$ as functions of n are shown in Table III cols. 1 and 2. The expected values u obtained from (6) are given in col. 3. For small samples, the two values $\tilde{x}_n$ and u differ widely, as might be expected. We are inclined to conclude that the asymptotic distribution of the range cannot hold for small samples. However, the only legitimate conclusion to be drawn is, that we cannot calculate the two parameters in the way stated before (6) and (8). Instead, we estimate them directly from the observations. The question of the most efficient estimates of these parameters is not yet solved. The simplest way is to use the mean range $\bar{w}_n$ and the standard deviation of the range $\sigma_{w,n}$ as given by Tippett [20] and Pearson [15]. To distinguish these estimates from the asymptotic values, we write the estimates with an index n. From (14) we obtain

$$(39) \qquad \frac{1}{\alpha_n} = \frac{\sqrt{3}}{\pi}\,\sigma_{w,n}\,; \qquad 2u_n = \bar{w}_n - \frac{2\gamma}{\alpha_n}\,.$$

Table III gives the calculated means w_n and standard deviations $\sigma_{w,n}$ of the range, and the estimates $1/\alpha_n$ and $2u_n$. Fig. 6 shows how the most probable

largest values $\bar{x}_n$ approach the expected largest value u with increasing sample size. The estimate u_n quickly approaches u. Besides we trace the mean range $\bar{w}_n$, the standard error of the range $\sigma_{w,n}$, and $1/\alpha_n$ which is proportional to it.
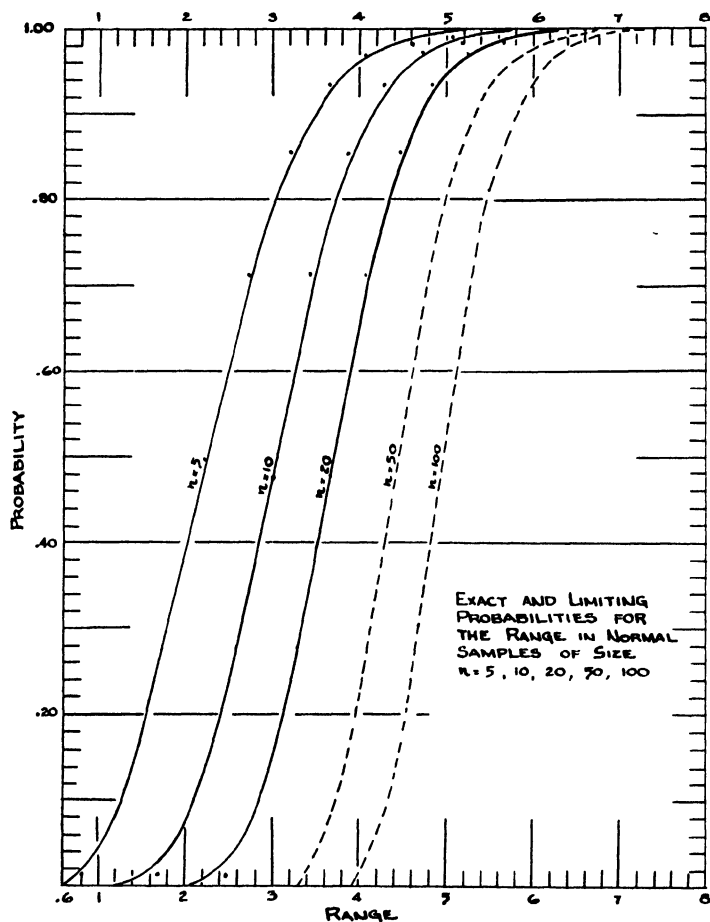
FIG. 4

From col. 8 follows that the condition $\alpha u \geqq 2$ is fulfilled from $n \geqq 6$ onward. The ranges obtained from the transformations

$$(40) \qquad\qquad w = 2u_n + \frac{R}{\alpha_n}$$

are given in Table IV, cols. 3–7. The asymptotic probabilities of the range as obtained from the combination of columns 3–7, and col. 2 of Table IV are traced

in Fig. 4 as separated points. The asymptotic probabilities are situated very near to the exact ones. Therefore the same method was used to calculate the asymptotic probabilities of the range for $n = 50$ and $n = 100$ which have not been calculated by Pearson. They too are traced in Fig. 4.
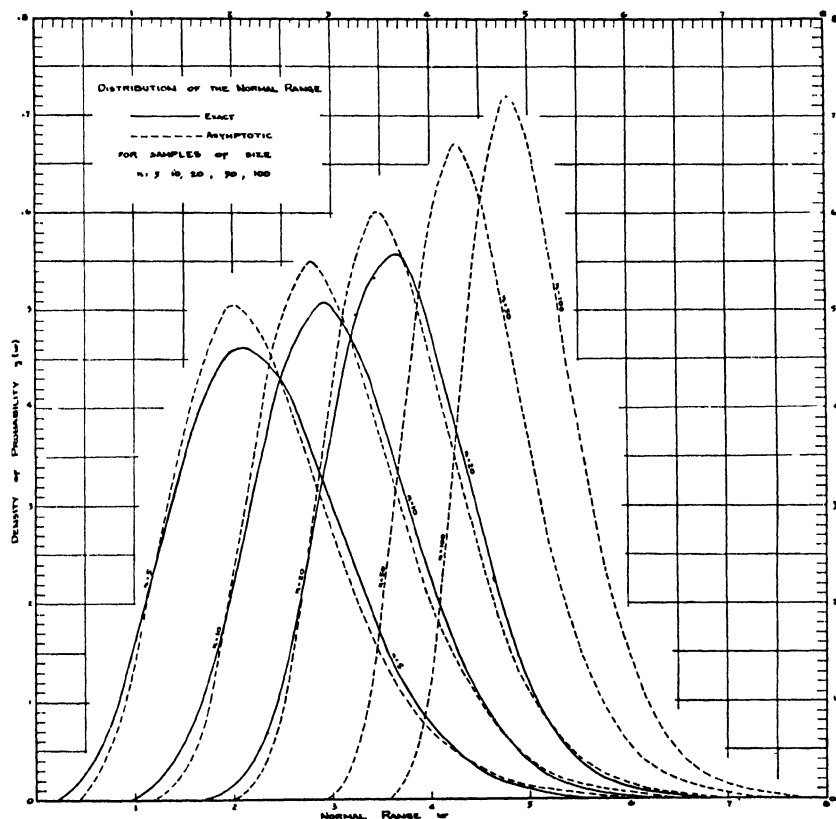
Fig. 5

The asymptotic probabilities of the range hold even for small normal samples. However, the parameters obtained from the exact distribution differ considerably from their asymptotic values. In other words: *The asymptotic probabilities of the range hold even for small normal samples provided that the parameters are taken from the observations.*

To compare the asymptotic distributions of the normal range to the calculated distributions, we attribute the asymptotic differences $\Delta\Psi/\alpha_n$ for a unit interval $\Delta w = 1$ to the middle of the corresponding intervals. The results are traced in Fig. 5 for $n = 5, 10, 20, 50, 100$. On the other hand, we take the differences

$\Delta\Psi$ for unit intervals from Pearson's tables, and trace them in the same graph. The fit of the calculated to the asymptotic values may be considered satisfactory.

TABLE III
Estimate of Parameters from the Calculated Distributions of the Normal Range

1	2	3	4	5	6	7	8
Sample size n	Largest Value		Mean Range $\bar{w}_n$	Standard deviation $\sigma_{w,n}$	Estimated parameters of the range		Lower limit $2\alpha_n u_n$
	Modal $\tilde{x}_n$	Expected u			$1/\alpha_n$	$2u_n$	
3	.765	.431	1.693	.8884	.4898	1.128	2.30
4	.938	.674	2.059	.8798	.4851	1.499	3.09
5	1.061	.842	2.326	.8641	.4764	1.776	3.73
10	1.419	1.282	3.078	.797	.439	2.571	5.86
20	1.740	1.645	3.735	.729	.402	3.271	8.14
50	2.126	2.054	4.498	.653	.360	4.082	11.34
100	2.377	2.326	5.015	.605	.334	4.630	13.86

TABLE IV
Asymptotic Probabilities for Normal Ranges Taken from Small Samples

1	2	3	4	5	6	7
Reduced range R	Probability $G(w) = \Psi(R)$	Normal ranges $w = 2u_n + R/\alpha_n$ for sample sizes				
		$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = 100$
-3	.000	.35	1.25	2.07	3.00	3.62
-2	.014	.82	1.69	2.47	3.36	3.96
-1	.093	1.30	2.13	2.87	3.72	4.30
0	.280	1.78	2.57	3.27	4.08	4.63
1	.517	2.52	3.01	3.67	4.44	4.96
2	.719	2.73	3.45	4.07	4.80	5.30
3	.853	3.21	3.89	4.48	5.16	5.63
4	.929	3.68	4.33	4.88	5.52	5.97
5	.967	4.16	4.77	5.28	5.88	6.30
6	.985	4.63	5.20	5.68	6.24	6.63
7	.994	5.11	5.64	6.09	6.60	6.97

Fig. 5 shows furthermore how the distributions of the range are shifted toward the right and become more concentrated for increasing sample sizes.

As an example for the practical application of the asymptotic distribution of the range, we use an observed distribution of 50 ranges taken from samples of

THE PARAMETERS OF THE
RANGE FOR SMALL
NORMAL SAMPLES

$\bar{w}_n$ MEAN RANGE

$\tilde{x}_n$ MODAL LARGEST VALUE

u SOLUTION OF EQUATION

u_n ESTIMATE OF u

σ_w STANDARD DEVIATION
OF THE RANGE

$\frac{1}{\sigma_n}$ PARAMETER

SAMPLE SIZE n

FIG. 6

$\frac{\bar{w}_n}{2}$ $\tilde{x}_n$

σ_w

u

u_n

$\frac{1}{\sigma_n^2}$

AVERAGES $\bar{u}_n$, u_n, $\frac{1}{2}\tilde{x}_n$, $\frac{1}{2}(\frac{1}{n})\bar{w}_n$ AND MEASURES OF DISPERSION σ_w, $\frac{1}{\sigma_n}$

407
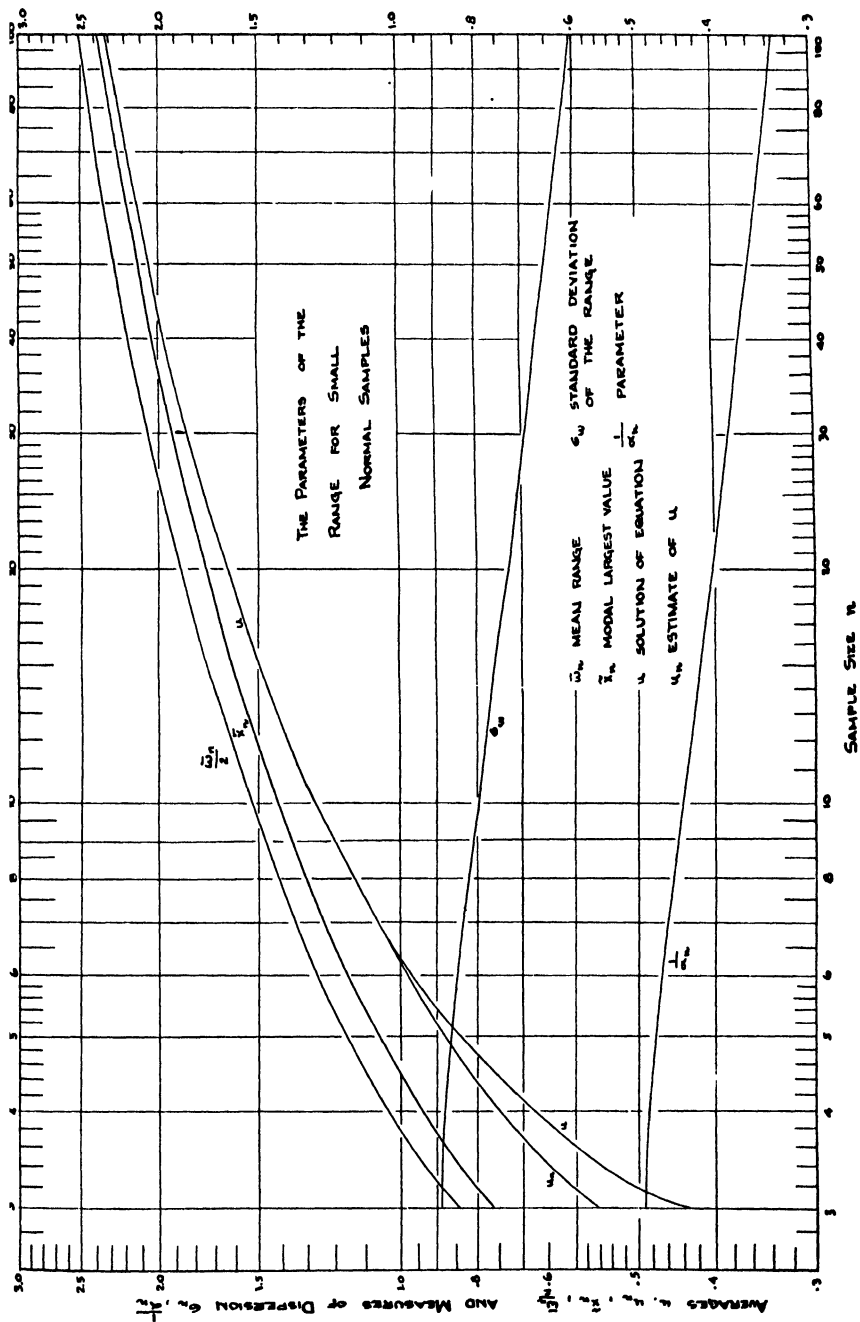
$n = 14$ normal values given in Freeman's book [5] p. 128. The observed step function is traced in Fig. 7. For reasons given in a previous article [7] we attribute the cumulative frequency .5 to the smallest range 3, and the cumulative frequency 49.5 to the largest range 18. To compare this step function with the
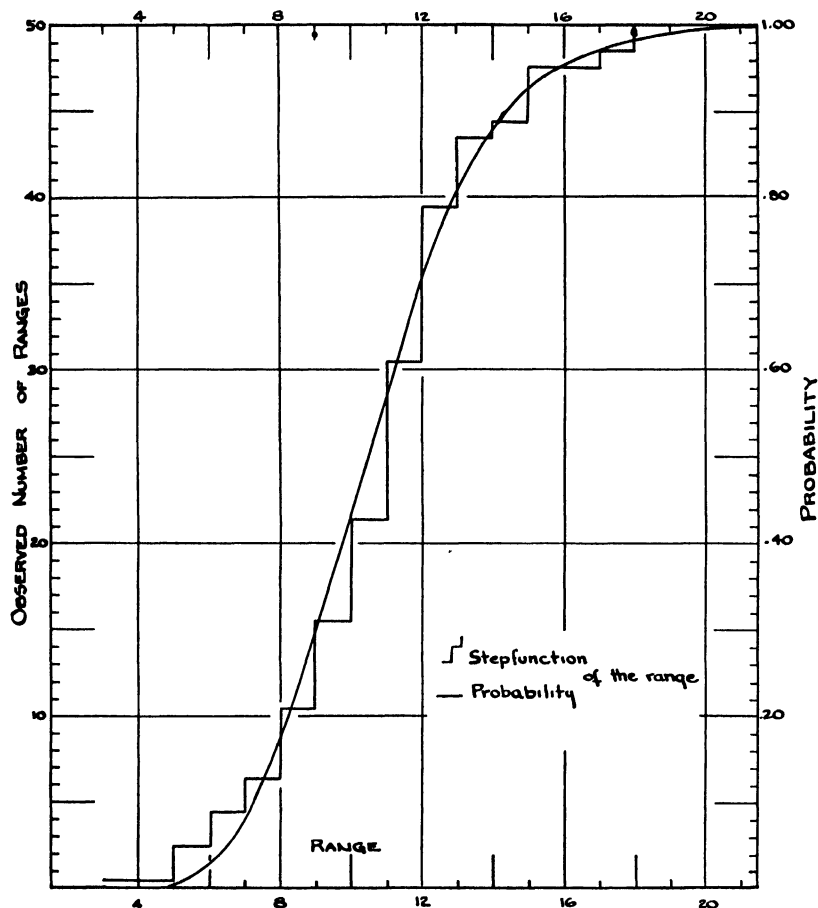
FIG. 7

probability $G(w)$, we estimate the two parameters u_n and α_n from formula (39). The mean range $\bar{w}_n$ and the estimate $s_{w,n}$ of the standard deviation of the ranges are

$$\bar{w} = 10.68; \qquad s_{w,n} = 2.93.$$

Consequently we obtain, from (39)

$$\frac{1}{\alpha} = 1.61; \qquad 2u_n = 8.82.$$

The theoretical ranges are thus, from (40),
$$w = 8.82 + 1.61\,R.$$
The corresponding probabilities $G(w)$ taken from Table I are traced in Fig. 7. The fit of the theory to the observations is certainly satisfactory, especially if we take into account that the ranges are given in integer numbers only.

8. The mth range and the asymmetrical case. An obvious generalization of the theory as established in paragraph 4 consists in the construction of the asymptotic distribution of the mth range for an unlimited symmetrical distribution of the exponential type. The mth range is the positive distance from the mth observation from above, x_m, to the mth observation from below, $_mx$. We suppose m to be very small compared to the sample size. Under the conditions stated in the beginning, the joint distribution $\mathfrak{w}_n(_mx,\ x_m)$ of the mth extreme values splits into the product of the asymptotic distribution of the mth extreme value from above, $f_m(x_m)$, by the asymptotic distribution of the mth extreme value from below, $_mf(_mx)$. Here, [6]

$$f_m(x_m) = \alpha_m \exp\left[-m\alpha_m(x_m - u_m) - me^{-\alpha_m(x_m - u_m)}\right]$$

$$_mf(_mx) = \alpha_m \exp\left[m\alpha_m(_mx + u_m) - me^{\alpha_m(_mx + u_m)}\right]$$

The sample size must be so large that the most probable mth extreme value $\tilde{x}_m$ is sufficiently near to u_m which is defined as the solution of

$$\Phi(u_m) = 1 - \frac{m}{n}.$$

The factor α_m defined by

$$\alpha_m = \frac{\varphi(u_m)}{1 - \Phi(u_m)}$$

is related to the asymptotic standard error σ_m of the mth extreme value by

$$\alpha_m \sigma_m = \sum_{\nu = m}^{\infty} \frac{1}{\nu^2}.$$

The joint asymptotic distribution $\mathfrak{w}(_mx,\ x_m)$ of the mth smallest value and the mth range

(41) $$w_m = x_m - {}_mx$$

is

$$\mathfrak{w}(_mx,\ w_m) = \alpha_m^2 \exp\left[-m\alpha_m(w_m - 2u_m) - me^{\alpha_m(_mx + u_m)} - me^{-\alpha_m(_mx + w_m - u_m)}\right].$$

The asymptotic distribution $g(w_m)$ of the mth range is, dropping the index m of the variable $_mx$,

$$g(w_m) = \alpha_m^2 e^{-m\alpha_m(w_m - 2u_m)} \int_{-\infty}^{+\infty} \exp\left[-me^{\alpha_m(x + u_m)} - me^{-\alpha_m(x + w_m - u_m)}\right] dx.$$

Again we introduce a reduced range R_m defined by

(42) $$\alpha_m(w_m - 2u_m) = R_m \geqq -2\alpha_m u_m$$

and put for the integration

$$\alpha_m(x + u_m) = y.$$

Then the asymptotic distribution $\psi(R_m)$ of the reduced mth range is

(43) $$\psi(R_m) = e^{-mR_m} \int_{-\infty}^{+\infty} \exp[-me^y - me^{-y-R_m}]\, dy.$$

The probability $\Psi(R_m)$ for the mth range

$$\Psi(R_m) = \int_{-2\alpha_m u_m}^{R_m} \psi(z)\, dz$$

cannot be reduced to a single integral. This is due to the fact that the probabilities of the mth extreme values cannot be written down except in the integral form [6]. No differential equation similar to (17) exists. However, the function (43) could be calculated by numerical methods. The mean $\bar{R}_m$, the generating function and the moments of the mth range have been given in a previous paper [8].

For sake of completeness, consider finally an unlimited asymmetrical initial distribution of the exponential type. In this case, the joint distribution of the smallest and of the largest value splits again, for large samples, into the product of the asymptotic distributions $f_1(x_1)$ and $f_n(x_n)$ of the smallest and of the largest values which are now [6]

$$f_1(x_1) = \alpha_1 \exp[\alpha_1(x_1 - u_1) - e^{\alpha_1(x_1-u_1)}];$$
$$f_n(x_n) = \alpha_n \exp[-\alpha_n(x_n - u_n) - e^{-\alpha_n(x_n-u_n)}].$$

Here, α_n and u_n are defined, as previously, by (6) and (8). The sample must be so large that the most probable smallest value $\tilde{x}_1$ is sufficiently near to the solution of

$$\Phi(u_1) = \frac{1}{n}.$$

The factor α_1 defined by

$$\alpha_1 = \frac{\varphi(u_1)}{\Phi(u_1)}$$

is related to the asymptotic standard error of the smallest value by

$$\alpha_1 \sigma_1 = \frac{\pi}{\sqrt{6}}.$$

The joint asymptotic distribution of the smallest value x_1 and the range w

$$\mathfrak{w}(x_1, w) = \alpha_1 \alpha_n \exp[\alpha_1(x_1 - u_1) - \alpha_n(x_1 + w - u_n) - e^{\alpha_1(x_1-u_1)} - e^{-\alpha_n(x_1+w-u_n)}]$$

contains four parameters instead of the two which exist in the symmetrical case. However, the number of parameters may be reduced to one. We introduce a reduced range R defined by

$$(44) \qquad R = \alpha_n(w - u_n + u_1)$$

being the range itself minus the range of the modes divided by a factor proportional to the standard error of the largest value. If we put

$$(45) \qquad \alpha_1(x_1 - u_1) = y; \qquad \frac{\alpha_n}{\alpha_1} = \beta$$

the distribution $\psi(R)$ of the reduced range becomes, in the asymmetrical case,

$$(46) \qquad \psi(R) = e^{-R} \int_{-\infty}^{+\infty} \exp[y(1 - \beta) - e^y - e^{-\beta y - R}]\, dy$$

and the probability $\Psi(R)$ for the reduced range is

$$(47) \qquad \Psi(R) = \int_{-\infty}^{+\infty} \exp[y - e^y - e^{-\beta y - R}]\, dy$$

a formula which may immediately be verified by differentiation with respect to R. The mode $\tilde{R}$ of the range is the solution of

$$\psi(\tilde{R}) = e^{-\tilde{R}} \int_{-\infty}^{+\infty} \exp[y(1 - 2\beta) - R - e^y - e^{-\beta y - R}]\, dy.$$

Contrary to the symmetrical case, the latter integral cannot be expressed by the probability, and no simple differential equation similar to (17) exists. The expressions (46) and (47) contain a single constant β measuring the asymmetry of the initial distribution. In the symmetrical case, $\beta = 1$, we obtain, of course, the previous formulas (12) and (13). In the asymmetrical case, the mean, the variance, and the higher moments of the mth range may be derived from the generating function given in a previous paper [8].

The asymptotic distribution of the mth range in the asymmetrical case can easily be obtained by combining the two procedures used in this paragraph.

REFERENCES

[1] L. von Bortkiewicz, "Variationsbreite und mittlerer Fehler," *Sitzungsberichte d. Berliner Math. Gesellschaft*, Vol. 21, (1921).

[2] ———, "Die Variationsbreite beim Gauss'schen Fehlergesetz," *Nordisk Statistisk Tidskrift*, Vol. 1 (1922).

[3] A. G. Carlton, "Estimating the parameters of a rectangular distribution," *Annals of Math. Stat.*, Vol. 17 (1946).

[4] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample", *Proc. Cambridge Phil. Soc.*, Vol. 24 (1928).

[5] H. A. Freeman, *Industrial Statistics*, John Wiley and Sons, 1942.

[6] E. J. Gumbel, "Les valeurs extrèmes des distributions statistiques", *Ann. Inst. H. Poincaré*, Vol. 4 (1935).

[7] ———, "On serial numbers", *Annals of Math. Stat.*, Vol. 14 (1943).

[8] ———, "Ranges and midranges", *Annals of Math. Stat.*, Vol. 15 (1944).

[9] ———, "On the independence of the extremes in a sample", *Annals of Math. Stat.*, Vol. 17 (1946).

[10] H. O. HARTLEY, "The range in random samples", *Biometrika*, Vol. 32 (1942).

[11] M. G. KENDALL, *The Advanced Theory of Statistics*, Vol. 1, London, 1943.

[12] A. T. McKAY AND E. S. PEARSON, "A note on the distribution of range in sample sizes of n", *Biometrika*, Vol. 25 (1933).

[13] ———, "Distribution of the difference between the extreme observations and the sample mean in samples of n from a normal universe", *Biometrika*, Vol. 27 (1935).

[14] BRITISH ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, *Mathematical Tables Vol. VI: Bessel Functions*, Part I: Functions of order zero and unity, Cambridge Univ. Press, 1937.

[15] E. S. PEARSON, "A further note on the distribution of range in samples taken from a normal population", *Biometrika*, Vol. 18 (1926).

[16] E. S. PEARSON AND H. O. HARTLEY, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942).

[17] KARL PEARSON, *Tables for Statisticians and Biometricians, Part II*, Cambridge Univ. Press, 1931.

[18] STUDENT, "Errors in routine analysis", *Biometrika*, Vol. 19 (1927).

[19] ARNOLD N. LOWAN (technical Director), *Table of the Bessel Functions $K_0(x)$ and $K_1(x)$ for x between zero and one*, Math. Tables Proj., New York.

[20] L. H. C. TIPPETT, "On the extreme individuals and the range of samples taken from a normal population", *Biometrika*, Vol. 17 (1925).

ADDITION AT PROOF READING:

G. Elfving's article "The asymptotical distribution of range in samples from a normal population", *Biometrika*, Vol. 35 (1947), appeared when this manuscript was ready for print. Elfving considers a probability transformation of the range whereas we deal with the range itself. His distribution requires the knowledge of the initial distribution and of the sample size, whereas this knowledge is not required in our asymptotic formula.

LOW MOMENTS FOR SMALL SAMPLES: A COMPARATIVE STUDY OF ORDER STATISTICS

By Cecil Hastings, Jr., Frederick Mosteller, John W. Tukey, and Charles P. Winsor

Douglas Aircraft Co., Harvard University, Princeton University, and The Johns Hopkins University

1. Summary. The means, variances, and covariances for samples of size ≤ 10 from the normal distribution, a selected long-tailed distribution, and the uniform distribution are tabled and compared with the usual asymptotic approximations. The methods of computation used and the accuracy expected are discussed. Use is made of the representation of an arbitrarily distributed variate as a monotone function of a uniformly (rectangularly) distributed variate. It is hoped that these tables will encourage experimentation with new statistical procedures.

2. Introduction. Two sorts of statistical procedures have been widely exploited in theoretical statistics—first the use of linear and quadratic combinations of the unordered observations and, second, the use of ranked (ordered) observations. Statistics based on ordered observations have recently been dubbed *systematic statistics* [2, Mosteller, 1946]. Analytic processes and a few necessary numerical tables have advanced the study of the first procedure greatly, at least for the special case of the normal distribution; but analytic procedures have not done much to exhibit the behavior of systematic statistics and the necessary tables have been lacking.

It would be very helpful to have (1) at least the first two moments (including product moments) of the order statistics, and (2) tables of the percentage points of their distributions, for samples of sizes from 1 to some moderately large value such as 100 and for a large representative family of distributions. This is a large order and will require much computation.

The first step in this direction was taken by Fisher and Yates [1] by tabulating the means, to two decimal places, of all order statistics from normal samples of size ≤ 50. The present paper continues the process by supplying all means, variances, and covariances for samples of size ≤ 10 from (a) the normal distribution, (b) the uniform (rectangular) distribution, (c) a special distribution with long tails. For purposes of comparison, we also supply approximate means, variances, and covariances for the uniform and the special distribution computed from suitable asymptotic formulas.

The special distribution has the representing function

$$(1) \qquad r(u) = (1 - u)^{-1/10} - u^{-1/10},$$

where u has the uniform distribution on the interval $[0, 1]$, and $x = r(u)$ is the variable whose order statistics interest us. This special distribution was es-

pecially constructed 1) to have high tails and 2) to provide moments of order statistics in closed form which could be evaluated with a reasonable amount of labor. The normal distribution is rather unreasonable in this latter respect—there being no known expression except in terms of single and double quadratures of some considerable numerical difficulty.

We have restricted ourselves to samples of size ≤ 10, and to only three distributions, all of these symmetrical, because of limited man-power rather than limited interest. Additional tables of a similar nature will surely prove helpful.

In order to obtain even as much information as provided in this paper, it has been necessary to make a joint effort, dividing the labor. The various parts of the work have been carried out more or less separately by the various authors—the means and variances for the normal by Mosteller, the covariances for the normal (which, with their double quadratures, required far more time than all the other thought and computation combined) by Hastings with some assistance from Mosteller, the choice of the special distribution by Tukey, and the computation for it by Winsor.

3. Results. In this section we provide the various tables that have been computed.

Table I gives the mean and standard deviation of the ith order statistic $x(i \mid n)$, [or $x_{i|n}$, we use whichever notation seems less likely to confuse and agree that $x(1 \mid n) \geq x(2 \mid n) \geq \cdots \geq x(n \mid n)$] from a sample of size n drawn from a uniform (U), normal (N), and a special distribution (S). All three distributions have been adjusted to have zero mean and unit variance. In addition Table I gives approximations for the mean and standard deviation as computed from asymptotic formulas for the normal (AN) and the special (AS).

If $f(x)$ is the density function, the asymptotic approximation for the mean $m(i \mid n)$ of the ith order statistic from a sample of size n is obtained by solving the equation

$$\int_{m(i \mid n)}^{\infty} f(x) \, dx = i/(n + 1)$$

for $m(i \mid n)$. Similarly the formula used for the asymptotic variance of $x(i \mid n)$ is

$$\frac{i(n - i + 1)}{n(n + 1)^2 \{f[m(i \mid n)]\}^2} \cdot$$

Values are given for $n = 1, 2, \cdots, 10$ and $i = 1, \cdots, \left[\dfrac{n}{2}\right]$. If $m(i \mid n)$ is an entry in the table for means, a missing entry $m(n - i + 1 \mid n) = -m(i \mid n)$; if $w(i \mid n)$ is an entry in the table of standard deviations, a missing entry

$$w(n - i + 1 \mid n) = w(i \mid n).$$

Table II gives the variances and covariances of the order statistics for the normal distribution (N) and the same quantities as approximated by the asymp-

TABLE I

*Means and standard deviations of order statistics $x(i|n)$ for uniform distribution
(U), normal (N), special (S), asymptotic normal (AN),
asymptotic special (AS)*

		Mean			Standard Deviation		
n	i	U AN	N AS	S	U AN	N AS	S
1	1	0 0	0 0	0	1.00000 1.2533	1.00000 .9804	1.00000
2	1	.57735 .4307	.56419 .3418	.53493	.81650 .9168	.82565 .7486	.84490
3	1	.86603 .6745	.84628 .5466	.80240	.67082 .7867	.74798 .6823	.82783
	2	0 0	0 0	0	.77460 .7236	.66983 .5660	.58457
4	1	1.03923 .8416	1.02938 .6954	.98473	.56569 .7144	.70122 .6542	.82982
	2	.34641 .2533	.29701 .1992	.25540	.69282 .6340	.60038 .5035	.52582
5	1	1.15470 .9674	1.16296 .8136	1.12449	.48795 .6670	.66898 .6415	.83642
	2	.57735 .4307	.49502 .3418	.42567	.61721 .5798	.55814 .4730	.50390
	3	0 0	0 0	0	.65465 .5605	.53557 .4384	.44903
6	1	1.23718 1.0676	1.26721 .9114	1.23847	.42857 .6331	.64492 .6330	.84423
	2	.74231 .5659	.64176 .4539	.55458	.55328 .5426	.52874 .4567	.49425
	3	.24744 .1800	.20155 .1412	.16785	.60609 .5147	.49620 .4057	.41648

TABLE I (*Continued*)

		Mean					Standard Deviation				
n	*i*	U	AN	N	AS	S	U	AN	N	AS	S
7	1	1.29904	1.1504	1.35218	.9957	1.33506	.38188	.6072	.62603	.6141	.85217
	2	.86603	.6745	.75737	.5462	.65892	.50000	.5150	.50670	.4359	.48992
	3	.43301	.3186	.35271	.2512	.29375	.55902	.4826	.46875	.3772	.39963
	4	0	0	0	0	0	.57735	.4737	.45874	.3617	.37747
8	1	1.34715	1.2207	1.42360	1.0697	1.41892	.31427	.5867	.61066	.6276	.85988
	2	.96225	.7647	.85222	.6259	.74690	.45542	.4936	.48930	.4402	.48823
	3	.57735	.4307	.47282	.3418	.39498	.51640	.4584	.44807	.3743	.38998
	4	.19245	.1397	.15251	.1094	.12502	.54433	.4447	.43264	.3494	.35616
9	1	1.38564	1.2816	1.48501	1.1358	1.49358	.31334	.5691	.59780	.6268	.86725
	2	1.03923	.8416	.93230	.6954	.82317	.41779	.4763	.47508	.4361	.48800
	3	.69282	.5244	.57197	.4191	.47995	.47863	.4393	.43171	.3722	.38414
	4	.34641	.2533	.27453	.1992	.22504	.51168	.4227	.41303	.3356	.34321
	5	0	0	0	0	0	.52223	.4178	.40751	.3268	.33173

TABLE I (*Concluded*)

		Mean				Standard Deviation		
n	i	U AN	N AS	S		U AN	N AS	S
10	1	1.41713 1.53875	1.56057			.28748 .58681	.87423	
		1.3352 1.1956				.5557 .6275		
	2	1.10222 1.00135	.89062			.38569 .46318	.48859	
		.9085 .7574				.4619 .4334		
	3	.78730 .65608	.55336			.44536 .41826	.38054	
		.6046 .4866				.4238 .3604		
	4	.47238 .37572	.30866			.48105 .39756	.33477	
		.3488 .2754				.4052 .3261		
	5	.15746 .12274	.09961			.49793 .38857	.31190	
		.1142 .0894				.3973 .3117		

totic formulas (AN). The asymptotic covariance between $x(i \mid n)$ and $x(j \mid n)$ is given by

$$\frac{j(n-i+1)}{n(n+1)^2 f[m(i\mid n)] f[m(j\mid n)]}, \qquad j \le i.$$

Symmetry relations exist for supplying the missing entries,

$$\text{cov}\ [x(i \mid n), x(j \mid n)] = \text{cov}\ [x(n-i+1\mid n), x(n-j+1\mid n)].$$

It might seem more natural to use the factor $n + 2$ rather than n in the denominator of the asymptotic variances and covariances so that the formulas would more nearly agree with those for the uniform distribution. However the use of n gives much better approximations for the normal and the special distribution.

Table III gives the variances and covariances of the order statistics for the uniform distribution (U), and Table IV gives the corresponding results for the special distribution (S). Table V gives the asymptotic variances and covariances for the special distribution (AS).

Table VI compares the correlation coefficients between the order statistics $x(i \mid n)$ and $x(j \mid n)$ for the uniform (U), the normal (N), and the special distribution (S).

It seems worthwhile to call attention to the following:

(1). Even for $n = 10$, the asymptotic formulas do not give satisfactory mean values for the order statistics.

(2). For $n \ge 8$, the asymptotic standard deviations for the normal are close

enough to be very useful.   For the special distribution we must except the two order statistics on each end from this statement.

TABLE II

Variances and covariances of the order statistics $x(i|n)$ for the normal (N) and the asymptotic normal (AN)

n	i	1 N	1 AN	2 N	2 AN	3 N	3 AN	4 N	4 AN	5 N	5 AN	6 N	6 AN	7 N	7 AN	8 N	8 AN	9 N	9 AN	10 N	10 AN
2	1	.68	.84	.32	.42																
3	1	.56	.62	.28	.33	.17	.21														
	2			.45	.52																
4	1	.49	.51	.24	.28	.16	.18	.11	.13												
	2			.36	.40	.24	.27														
5	1	.45	.44	.22	.24	.15	.17	.11	.12	.07	.09										
	2			.31	.34	.21	.23	.15	.17												
	3					.29	.31														
6	1	.42	.40	.21	.22	.13	.15	.11	.12	.07	.09	.06	.07								
	2			.28	.29	.19	.20	.14	.15	.10	.12										
	3					.25	.26	.18	.20												
7	1	.39	.37	.20	.20	.13	.14	.10	.11	.08	.09	.06	.07	.05	.05						
	2			.26	.27	.17	.19	.13	.14	.10	.11	.08	.09								
	3					.22	.23	.17	.18	.13	.14										
	4							.21	.22												
8	1	.37	.34	.19	.19	.13	.13	.09	.10	.08	.08	.06	.07	.04	.05	.04	.04				
	2			.24	.24	.17	.17	.12	.13	.10	.10	.08	.09	.07	.07						
	3					.20	.21	.15	.16	.12	.13	.09	.11								
	4							.19	.20	.15	.16										
9	1	.36	.32	.18	.18	.12	.13	.09	.10	.07	.08	.06	.07	.05	.05	.04	.05	.04	.04		
	2			.23	.23	.16	.16	.11	.12	.09	.10	.08	.08	.06	.07	.05	.06				
	3					.19	.19	.14	.15	.11	.12	.10	.10	.08	.08						
	4							.17	.18	.14	.14	.12	.12								
	5									.17	.17										
10	1	.34	.31	.17	.17	.12	.12	.09	.09	.07	.08	.06	.06	.05	.05	.04	.05	.03	.04	.03	.03
	2			.21	.21	.14	.15	.11	.12	.09	.09	.07	.08	.06	.07	.05	.06	.04	.05		
	3					.17	.18	.13	.14	.11	.11	.09	.09	.08	.08	.06	.07				
	4							.16	.16	.12	.13	.11	.11	.09	.09						
	5						.			.15	.16	.13	.13								

(3). For $n \geq 8$, the asymptotic variances and covariances of the normal are close enough for many, if not most purposes.

(4). For the special distribution, only the variances and covariances of moderately central order statistics are adequately given by the asymptotic formulas.

TABLE III

Variances and covariances for the uniform distribution (U)

n	i \ j	1	2	3	4	5	6	7	8	9	10
2	1	.66667	.33333								
3	1	.45000	.30000	.15000							
	2		.60000								
4	1	.32000	.24000	.16000	.08000						
	2		.48000	.32000							
5	1	.23810	.19047	.14286	.09522	.04762					
	2		.38095	.28571	.19047						
	3			.42857							
6	1	.18367	.15306	.12245	.09184	.06122	.03061				
	2		.30612	.24490	.18367	.12245					
	3			.36735	.27551						
7	1	.14583	.12500	.10417	.08333	.06250	.04167	.02083			
	2		.25000	.20833	.16667	.12500	.08333				
	3			.31250	.25000	.18750					
	4				.33333						
8	1	.11852	.10370	.08889	.07407	.05925	.04444	.02963	.01481		
	2		.20741	.17778	.14815	.11852	.08889	.05925			
	3			.26667	.22222	.17778	.13333				
	4				.29630	.23704					
9	1	.09818	.08727	.07636	.06545	.05455	.04363	.03273	.02182	.01091	
	2		.17455	.15273	.13091	.10909	.08727	.06545	.04363		
	3			.22909	.19636	.16364	.13091	.09818			
	4				.26182	.21818	.17455				
	5					.27273					
10	1	.08264	.07438	.06611	.05785	.04959	.04132	.03306	.02479	.01653	.00826
	2		.14876	.13223	.11570	.09917	.08264	.06611	.04959	.03306	
	3			.19835	.17355	.14876	.12397	.09917	.07438		
	4				.23140	.19835	.16529	.13223			
	5					.24793	.20661				

(5). The correlation coefficients change rather little from distribution to distribution, the poorest approximation being for end order statistics.

TABLE IV

Variances and covariances for the special distribution (S)

n	i	1	2	3	4	5	6	7	8	9	10
2	1	.71385	.28615								
3	1	.68530	.24214	.15957							
	2		.34172								
4	1	.68860	.23277	.14141	.11123						
	2		.27649	.17532							
5	1	.69960	.23154	.13655	.10004	.08614					
	2		.25391	.15418	.11490						
	3			.20163							
6	1	.71272	.23310	.13544	.09667	.07786	.07080				
	2		.24429	.14506	.10486	.08514					
	3			.17345	.12762						
7	1	.72619	.23577	.13582	.09565	.07517	.06109	.06012			
	2		.24002	.14065	.10004	.07913	.06776				
	3			.15970	.11509	.09184					
	4				.14249						
8	1	.73940	.23890	.13687	.09562	.07420	.06179	.05471	.05291		
	2		.23837	.13850	.09754	.07608	.06359	.05615			
	3			.15208	.10822	.08499	.07138				
	4				.12685	.10053					
9	1	.75211	.24219	.13825	.09608	.07398	.06085	.05266	.04789	.04721	
	2		.23814	.13756	.09625	.07143	.06141	.05327	.04852		
	3			.14756	.10413	.08097	.06707	.05835			
	4				.11780	.09225	.07680				
	5					.11004					
10	1	.76428	.24550	.13978	.09680	.07414	.06053	.05176	.04604	.04271	.04272
	2		.23872	.13732	.09565	.07354	.06018	.05156	.04594	.04266	
	3			.14481	.10158	.07846	.06444	.05533	.04940		
	4				.11207	.08707	.07180	.06186			
	5					.10016	.08300				

4. Methods of calculation and accuracy for the normal distribution. The means and variances of the order statistics for the normal distribution were obtained from direct quadrature of forms like

$$\int_{-\infty}^{\infty} x^k [F(x)]^j [1 - F(x)]^{n-j-1} f(x)\, dx, \qquad k = 1, 2,$$

where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt \quad \text{and} \quad f(x) = F'(x).$$

It is believed that the means are correct to within one unit in the fifth decimal and that the standard deviations are correct to within 2 or 3 units in the fifth decimal.

TABLE V

Variances and covariances of the special distribution as computed from asymptotic formulas

n	i \ j	1	2	3	4	5	6	7	8	9	10
2	1	.56044	.28022								
3	1	.46550	.22297	.15517							
	2		.32038								
4	1	.42792	.20168	.13444	.10698						
	2		.25347	.16898							
5	1	.41156	.19167	.12579	.09605	.08231					
	2		.22368	.14679	.11208						
	3			.19221							
6	1	.40072	.18667	.12105	.09080	.07464	.06679				
	2		.20861	.13529	.10147	.08341					
	3			.16457	.12343						
7	1	.37715	.17527	.11304	.08394	.06782	.05842	.05388			
	2		.19004	.12258	.09103	.07354	.06335				
	3			.14232	.10569	.08538					
	4				.13731						
8	1	.39389	.18276	.11746	.08669	.06935	.05873	.05221	.04924		
	2		.19382	.12458	.09194	.07355	.06229	.05538			
	3			.14011	.10341	.08272	.07005				
	4				.12211	.09769					
9	1	.39286	.18226	.11881	.08591	.06829	.05727	.05092	.04556	.04367	
	2		.19019	.12398	.08965	.07126	.05977	.05313	.04754		
	3			.13855	.10019	.07963	.06678	.05938			
	4				.11265	.08958	.07512				
	5					.10680					
10	1	.39373	.18242	.11677	.08560	.06775	.05646	.04891	.04379	.04054	.03937
	2		.18784	.12024	.08813	.06977	.05814	.05036	.04508	.04174	
	3			.12988	.09520	.07536	.06280	.05440	.04871		
	4				.10633	.08417	.07014	.06076			
	5					.09716	.08098				

The evaluation of the covariances was much more troublesome, requiring the evaluation of iterated integrals of the form

$$\int_{-\infty}^{\infty} xf(x)F^j(x) \int_{-\infty}^{x} tf(t)[1-F(t)]^i \, dt \, dx.$$

Necessary linear combinations of such forms give rise to considerable loss of accuracy. The covariances are believed to be correct to within 1 unit in the second decimal (except for one or two values which may be off by two units).

TABLE VI

Correlation coefficients × 10² between order statistics x(i | n), x(j | n) for the uniform (U), normal (N), and special distribution (S)

n	i	2 U	2 N	2 S	3 U	3 N	3 S	4 U	4 N	4 S	5 U	5 N	5 S	6 U	6 N	6 S	7 U	7 N	7 S	8 U	8 N	8 S	9 U	9 N	9 S	10 U	10 N	10 S
2	1	50	47	40																								
3	1	58	55	50	33	29	23																					
4	1	61	58	53	41	38	32	25	21	16																		
	2				67	65	63																					
5	1	63	60	55	45	41	36	32	28	24	20	16	12															
	2				71	70	68	50	48	42																		
6	1	65	61	56	47	44	39	35	32	28	26	23	19	17	13	10												
	2				73	72	70	55	53	51	40	38	35															
	3							75	74	74																		
7	1	66	62	56	49	45	40	38	34	30	29	26	22	22	19	15	14	11	08									
	2				74	73	72	58	56	54	45	43	40	33	31	28												
	3							78	77	76	60	59	58															
8	1	66	62	57	50	46	41	40	36	32	32	28	24	25	22	18	19	16	13	12	10	07						
	2				76	74	73	60	58	56	48	46	44	38	36	33	29	26	24									
	3							79	79	78	63	62	61	50	49	47												
	4										80	80	79															
9	1	67	63	57	51	47	42	41	37	32	33	30	26	27	24	20	22	19	16	17	14	11	11	09	06			
	2				76	75	73	61	60	57	50	48	46	41	39	37	33	31	28	25	23	20						
	3							80	80	79	66	65	64	54	52	51	43	41	40									
	4										82	81	81	67	66	65												
10	1	67	63	57	52	47	42	42	38	33	35	31	27	29	25	22	24	21	18	19	17	14	15	13	10	10	08	06
	2				77	76	74	62	61	58	52	50	48	43	42	40	36	34	32	29	27	25	22	20	18			
	3							81	80	80	67	66	66	56	55	54	46	45	43	38	35	34						
	4										83	82	82	69	69	68	57	56	55									
	5													83	83	83												

Better tables of these covariances are badly needed, and it is hoped that someone will provide them.

The asymptotic values are correct to the two decimals given.

5. Computation in terms of the representing function. It will prove convenient in working with the special distribution, as indeed it does in many statistical procedures, to introduce the *representing function* $r(u)$, which is a monotone function such that

$$Pr\{r(u_1) \leq x \leq r(u_2)\} = u_2 - u_1, \, u_2 \geq u_1 \, .$$

Thus if u has a uniform (= rectangular on $[0, 1]$) distribution then $x = r(u)$ defines a variate with the given distribution.

The ith order statistic of n from the uniform distribution, $u_{i|n}$, is distributed according to

$$i\binom{n}{i}u^{n-i}(1 - u)^{i-1} \, du, \, 0 \leq u \leq 1,$$

where it is important to remember that $u_{1|n}$ is the largest and not the smallest order statistic; and the joint distribution of $u = u_{i|n}$ and $v = u_{j|n}$, $(j > i)$, is given by

$$i(j - i)\begin{bmatrix} n \\ i, j - i, n - j \end{bmatrix}v^{n-j}(u - v)^{j-i-1}(1 - u)^{i-1} \, du \, dv, \quad 0 \leq v \leq u \leq 1,$$

where $\begin{bmatrix} n \\ i, j - i, n - j \end{bmatrix}$ is a multinomial coefficient.

The means, variances, and covariances which we desire can be written as follows (it is immaterial whether we think of expectations over x's or over u's):

$$E(x_{i|n}) = E(r(u_{i|n})) = i\binom{n}{i}\int_0^1 r(u)u^{n-i}(1 - u)^{i-1} \, du,$$

$$\text{var}\,(x_{i|n}) = E(x_{i|n})^2 - (E(x_{i|n}))^2 = E(r^2(u_{i|n})) - E(x_{i|n})^2$$

$$= i\binom{n}{i}\int_0^1 r^2(u)u^{n-i}(1 - u)^{i-1} \, du - (E(x_{i|n}))^2,$$

$$\text{cov}\,(x_{i|n}, x_{j|n}) = E(x_{i|n}x_{j|n}) - E(x_{i|n}) \cdot E(x_{j|n})$$

$$= E(r(u_{i|n})r(u_{j|n})) - E(x_{i|n})E(x_{j|n})$$

$$= i(j - i)\begin{bmatrix} n \\ i, j - i, n - j \end{bmatrix}\int_0^1\int_v^1 r(u)r(v)v^{n-j}(u - v)^{j-i-1}(1 - u)^{i-1} \, du \, dv$$

$$- E(x_{i|n})E(x_{j|n})$$

Introducing $E_{s,t}$ by

$$E_{s,t} = \int_0^1\int_v^1 r(u)r(v)u^s v^t \, du \, dv,$$

we have

$$E(x_{i|n}x_{j|n}) = i(j - 1)!\begin{bmatrix} n \\ i, j - i, n - j \end{bmatrix}$$

$$\cdot \sum_{k,m} (-1)^{k+m}\binom{j - i - 1}{k}\binom{i - 1}{m} E_{k+m, n-i-1-k},$$

and, in particular,

$$E(x_{1|2}x_{2|2}) = 2E_{0,0},$$

$$E(x_{1|5}x_{4|5}) = 60E_{2,1} - 120E_{1,8} + 60E_{0,8}.$$

Introducing $E_{s,s}$ by

$$E_{s,s} = \int_0^1 r^2(u)u^s \, du,$$

we have

$$E(x_{i|n}^2) = i\binom{n}{i} \sum_k (-1)^k \binom{i-1}{k} E_{n-i+k, n-i+k}$$

and, in particular,

$$E(x_{2|5}^2) = 20E_{3,3} - 20E_{4,4}.$$

Introducing E_s by

$$E_s = \int_0^1 r(u)u^s \, du,$$

·we have

$$E(x_{i|n}) = i\binom{n}{i} \sum_k (-1)^k \binom{i-1}{k} E_{n-i+k}$$

and, in particular,

$$E(x_{3|5}) = 30E_2 - 60E_3 + 30E_4.$$

Thus the computation of the desired means, variances, and covariances is reduced to the computation of the integrals E_s, $E_{s,s}$, and $E_{s,t}$.

We shall also want to calculate the asymptotic approximations to the means, variances, and covariances of the order statistics. For the uniform distribution, it is well known that

$$\text{mean } (u_{i|n}) = \frac{n-i+1}{n+1},$$

$$\text{var } (u_{i|n}) = \frac{i(n-i+1)}{(n+1)^2(n+2)},$$

$$\text{cov } (u_{i|n}u_{j|n}) = \frac{i(n-j+1)}{(n+1)^2(n+2)}, \qquad (i < j).$$

These asymptotic formulas are transformed from u to x by the relations $x = r(u)$ and $dx = r'(u) \, du$, giving

$$\text{approx mean } (x_{i|n}) = r\left(\frac{n-i+1}{n+1}\right),$$

$$\text{approx var } (x_{i|n}) = \left(r'\left(\frac{n-i+1}{n+1}\right)\right)^2 \frac{i(n-i+1)}{(n+1)^2(n+2)},$$

approx cov $(x_{i|n} x_{j|n}) = r' \left(\dfrac{n - j + 1}{n + 1} \right)$

$$\cdot \; r' \left(\dfrac{n - i + 1}{n + 1} \right) \dfrac{i(n - j + 1)}{(n + 1)^2(n + 2)}, \qquad (i \leq j),$$

as noted above, in our calculations we have replaced $n + 2$ by n in the denominator.

6. Reduction of integrals for the special case. When the representing function is

$$x = r(u) = \dfrac{1}{(1 - u)^\lambda} - \dfrac{1}{u^\lambda}, \qquad (\lambda > 0),$$

we obtain a symmetrical distribution with long tails. (For the normal distribution $r(u) = o(\ln u)$ as $u \to 0$). The integrals we want are

$$E_s = \int_0^1 \{(1 - u)^{-\lambda} - u^{-\lambda}\} u^s \, du,$$

$$E_{s,s} = \int_0^1 \{(1 - u)^{-\lambda} - u^{-\lambda}\}^2 u^s \, du,$$

$$E_{s,t} = \int_0^1 \int_v^1 \{(1 - u)^{-\lambda} - u^{-\lambda}\}\{(1 - v)^{-\lambda} - v^{-\lambda}\} u^s v^t \, du \, dv,$$

which can be expressed as

$$E_s \;\; = A_s(\lambda) - B_s(\lambda),$$

$$E_{s,s} = A_{s,s}(\lambda) - 2B_{s,s}(\lambda) + C_{s,s}(\lambda),$$

$$E_{s,t} = A_{s,t}(\lambda) - B_{s,t}(\lambda) - C_{s,t}(\lambda) + D_{s,t}(\lambda),$$

where

$$A_s(\lambda) = \int_0^1 (1 - u)^{-\lambda} u^s \, du = b(-\lambda, s),$$

$$B_s(\lambda) = \int_0^1 u^{-\lambda} u^s \, du = \dfrac{1}{s + 1 - \lambda},$$

$$A_{s,s}(\lambda) = \int_0^1 (1 - u)^{-2\lambda} u^s \, du = b(-2\lambda, s),$$

$$B_{s,s}(\lambda) = \int_0^1 (1 - u)^{-\lambda} u^{-\lambda} u^s \, du = b(-\lambda, s - \lambda),$$

$$C_{s,s}(\lambda) = \int_0^1 u^{-2\lambda} u^s \, du = \dfrac{1}{s + 1 - 2\lambda},$$

$$A_{s,t}(\lambda) = \int_0^1 \int_v^1 (1 - u)^{-\lambda}(1 - v)^{-\lambda} u^s v^t \, du \, dv$$

$$= \sum_{i=0}^{s} \binom{s}{i} (-)^i \frac{b(i + 1 - 2\lambda, t)}{i + 1 - \lambda},$$

$$B_{s,t}(\lambda) = \int_0^1 \int_v^1 (1 - u)^{-\lambda} v^{-\lambda} u^s v^t \, du \, dv$$

$$= \sum_{i=0}^{s} \binom{s}{i} (-)^i \frac{b(i + 1 - \lambda, t - \lambda)}{i + 1 - \lambda} = \frac{b(s + t + 1 - \lambda, -\lambda)}{t + 1 - \lambda},$$

$$C_{s,t}(\lambda) = \int_0^1 \int_v^1 u^{-\lambda} (1 - v)^{-\lambda} u^s v^t \, du \, dv$$

$$= \frac{1}{s + 1 - \lambda} \{ b(-\lambda, t) - b(s + t + 1, -\lambda, -\lambda) \},$$

$$D_{s,t}(\lambda) = \int_0^1 \int_v^1 u^{-\lambda} v^{-\lambda} u^s v^t \, du \, dv$$

$$= \frac{1}{(t + 1 - \lambda)(s + t + 2 - 2\lambda)},$$

where throughout

$$b(p, q) = \frac{p! \, q!}{(p + q + 1)!} = \frac{\Gamma(p + 1)\Gamma(q + 1)}{\Gamma(p + q + 2)} = B(p + 1, q + 1).$$

7. Calculations for the special distribution. The computations for the special distribution were made from the formulas in the preceding section.. The quantities A, B, C, D were computed from $r = s = 0$ to $r + s = 8$, whence the values of E_s, E_{ss}, E_{st} were calculated. The values of the means, variances, and covariances were then obtained from the formulas of section 3.

The means, variances, and covariances are believed to be accurate to the five decimal places given.

8. Formulas and accuracy for the uniform. The means, variances, and covariances of the uniform are given near the end of section 5. Since $r(u) \equiv u$, they are also the values given by the asymptotic approximation, when $n + 2$ is used.

The tabulated values were computed to six decimal places and rounded to the four or five decimals given.

REFERENCES

[1] R. A. FISHER AND F. YATES, *Statistical Tables*, Oliver and Boyd, London, 1943.
[2] FREDERICK MOSTELLER, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

SEQUENTIAL CONFIDENCE INTERVALS FOR THE MEAN OF A NORMAL DISTRIBUTION WITH KNOWN VARIANCE

By Charles Stein and Abraham Wald

Columbia University

1. Summary. We consider sequential procedures for obtaining confidence intervals of prescribed length and confidence coefficient for the mean of a normal distribution with known variance. A procedure achieving these aims is called optimum if it minimizes the least upper bound (with respect to the mean) of the expected number of observations. The result proved is that the usual non-sequential procedure is optimum.

2. Introduction. The problem of sequential confidence sets in general has been considered briefly by one of the authors [1]. Let $\{X_i\}$, $(i = 1, 2, \cdots)$, be a sequence of random variables whose distribution is specified except for the value of a parameter θ whose range is a space Ω. Sequential confidence sets are determined by a rule as to when to stop sampling, together with a function of the sample whose value is one of a specified class of subsets of Ω. The class of subsets is chosen in advance depending on the purpose of the estimation. For example, it may be the class of all intervals of prescribed length or the class of all sets whose diameter does not exceed a given value. It is required that the probability that this (random) set covers θ should be greater than or equal to a specified confidence coefficient α for all θ. A procedure for finding sequential confidence intervals is considered optimum if it minimizes some specified function of the expected numbers of observations. Here this function is taken to be the least upper bound. In contrast with the result of this paper, a case where sequential confidence intervals may have an advantage over non-sequential procedures has been given by one of the authors [2]. The X_i are independently normally distributed with unknown mean and unknown variance, and the problem is to find confidence intervals of fixed length for the unknown mean. As was first shown by Dantzig [3] this cannot be accomplished by a non-sequential procedure. Another case where this is true is the problem of finding confidence intervals of the form (p_0, kp_0) where k is a specified number greater than 1, for the probability in a binomial distribution.

Let $\{X_i\}$, $(i = 1, 2, \cdots)$, be independently normally distributed with unknown mean ξ and known variance σ_1^2. It is desired to specify a sequential procedure for obtaining confidence intervals of fixed length l for the mean ξ. This is provided by a rule according to which at each stage of the experiment, after obtaining the first m observations $X_1, \cdots, X_m$ for each integral value m, one makes one of the following decisions:

a) Take an $(m + 1)$st observation.

b) Terminate the procedure and state that the mean lies in the interval

$(Y - \frac{1}{2}l, Y + \frac{1}{2}l)$, where $Y = \mathfrak{C}_m(X_1, \cdots, X_m)$, $\mathfrak{C}_m$ being a measurable real-valued function. The serial number m of the observation on which the procedure terminates is, of course, a random variable and will be denoted by n.

For any relation R the symbol $P(R \mid \xi)$ will denote the probability that R holds when ξ is the true mean of X_i. The confidence coefficient of a sequential procedure S is defined by

$$(1) \qquad \alpha(S) = \underset{\xi}{g.l.b.} \; P(Y - \tfrac{1}{2}l < \xi < Y + \tfrac{1}{2}l \mid \xi).$$

Denote by $n_0(S)$ the maximum expected number of observations, i.e.

$$(2) \qquad n_0(S) = \underset{\xi}{l.u.b.} \; E(n \mid \xi, S)$$

where $E(n \mid \xi, S)$ denotes the expected value of n when ξ is the true mean and the procedure S is used.

A procedure S will be considered optimum if, for all S' such that $\alpha(S') = \alpha(S)$,

$$(3) \qquad n_0(S) \leq n_0(S').$$

It will be shown that an optimum procedure $S(\nu, c)$ can be obtained as follows:

a) For all $m < \nu$, a fixed positive integer, take another observation.

b) For $m = \nu$, terminate the procedure if

$$(4) \qquad \sum_1^\nu X_i^2 - \frac{1}{\nu}\left(\sum_1^\nu X_i\right)^2 > c\sigma_1^2$$

and let $Y = \dfrac{1}{\nu}\overset{\nu}{\underset{1}{\Sigma}}X_i$. (The inequality (4) is used merely as a device for fixing the probability of taking ν observations, this random event to be independent of whether $(Y - \frac{1}{2}l, Y + \frac{1}{2}l)$ covers ξ, given ν.)

c) Otherwise take a $(\nu + 1)$st observation, terminating the process, and let

$$Y = \frac{1}{\nu + 1} \sum_1^{\nu+1} X_i,$$

When $c = 0$, this is the usual non-sequential procedure.

Clearly,

$$(5) \qquad \alpha[S(\nu, c)] = P\{\chi_{\nu-1}^2 > c\}H\left(\frac{\sqrt{\nu}\,l}{2\sigma_1}\right) + [1 - P\{\chi_{\nu-1}^2 > c\}]H\left(\frac{\sqrt{\nu+1}\,l}{2\sigma_1}\right),$$

where

$$(6) \qquad H(u) = \frac{1}{\sqrt{2\pi}} \int_{-u}^{u} e^{-\frac{1}{2}x^2}\,dx = \sqrt{\frac{2}{\pi}} \int_0^u e^{-\frac{1}{2}x^2}\,dx.$$

Also

$$(7) \qquad n_0[S(\nu, c)] = \nu + 1 - P\{\chi_{\nu-1}^2 > c\},$$

By a proper choice of ν and c we can achieve any desired confidence coefficient

$\alpha \geq H\left(\dfrac{l}{\sqrt{2}\sigma_1}\right)$. There is no essential loss of generality in considering only the case $\sigma_1 = 1$, and this will be done in the remainder of this paper.

3. A lower bound for $n_0(S)$ and an upper bound for $\alpha(S)$. Consider any sequential procedure S for obtaining confidence intervals of length l. Put

(8) $$\alpha(\xi, S) = P\{Y - \tfrac{1}{2}l < \xi < Y + \tfrac{1}{2}l \mid \xi\}.$$

That is, $\alpha(\xi, S)$ is the probability that the confidence interval will cover the true mean ξ when the procedure S is used. According to (1),

(9) $$\alpha(S) = \operatorname*{g.l.b.}_{\xi} \alpha(\xi, S).$$

In order to obtain a lower bound for $n_0(S)$ and an upper bound for $\alpha(S)$, we suppose that the procedure S is applied when ξ is not a fixed number but a random variable normally distributed with mean 0 and variance σ^2. Then the probability that the confidence interval covers ξ is

(10) $$\bar{\alpha}(\sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} \alpha(\xi, S)\, d\xi \geq \alpha(S)$$

and the expected number of observations is

(11) $$\bar{E}(n \mid \sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} E(n \mid \xi, S)\, d\xi \leq n_0(S).$$

Let $p_m(\xi, S)$, $(m = 1, 2, \cdots, \text{ad. inf.})$, denote the probability that $n = m$ when ξ is the true mean and procedure S is used. Put

(12) $$\bar{p}_m(\sigma, S) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\xi^2/2\sigma^2} p_m(\xi, S)\, d\xi.$$

Since

(13) $$\bar{E}(n \mid \sigma, S) = \sum_{m=1}^{\infty} m\bar{p}_m(\sigma, S)$$

we obtain from (11)

(14) $$\sum_{m=1}^{\infty} m\bar{p}_m(\sigma, S) \leq n_0(S).$$

We shall now derive an upper bound for $\bar{\alpha}(\sigma, S)$. Since $X_i = \xi + \epsilon_i$ where the ϵ_i are independently normally distributed with mean 0 and variance 1, the joint distribution of ξ and X_i, $(i = 1, \cdots, m)$, is a multivariate normal distribution with

(15) $$E\xi = EX_i = 0$$

and covariance matrix

$$(16) \quad E \begin{pmatrix} \xi \\ X_1 \\ \vdots \\ X_m \end{pmatrix} (\xi, X_1, \cdots, X_m) = \begin{pmatrix} \sigma^2 & \sigma^2 & \cdots & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2+1 & \sigma^2 & \cdots & \sigma^2 \\ \vdots & & \sigma^2 & \sigma^2+1 & \cdots & \sigma^2 \\ \vdots & \vdots & & & & \vdots \\ \sigma^2 & \sigma^2 & \cdots & \cdots & \sigma^2+1 \end{pmatrix}$$

Thus the conditional distribution of ξ given $X_1, \cdots, X_m$ is normal with mean

$$E(\xi \mid X_1, \cdots, X_m) = (\sigma^2, \cdots, \sigma^2) \begin{pmatrix} \sigma^2+1 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2+1 & \cdots & \sigma^2 \\ \vdots & \vdots & & \vdots \\ \sigma^2 & \sigma^2 & \cdots & \sigma^2+1 \end{pmatrix}^{-1} \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$$

$$(17) \quad = \sigma^2(1, 1, \cdots, 1) \begin{pmatrix} \dfrac{(m-1)\sigma^2+1}{m\sigma^2+1} & -\dfrac{\sigma^2}{m\sigma^2+1} & \cdots & -\dfrac{\sigma^2}{m\sigma^2+1} \\ -\dfrac{\sigma^2}{m\sigma^2+1} & \dfrac{(m-1)\sigma^2+1}{m\sigma^2+1} & \cdots & -\dfrac{\sigma^2}{m\sigma^2+1} \\ \vdots & \vdots & & \vdots \\ -\dfrac{\sigma^2}{m\sigma^2+1} & -\dfrac{\sigma^2}{m\sigma^2+1} & \cdots & \dfrac{(m-1)\sigma^2+1}{m\sigma^2+1} \end{pmatrix}$$

$$\times \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} = \frac{\sigma^2}{m\sigma^2+1} \sum_1^m X_i$$

and variance

$$(18) \quad \sigma^2 - \frac{\sigma^4}{(m\sigma^2+1)^2} E \left(\sum_{i=1}^m X_i \right)^2 = \frac{\sigma^2}{m\sigma^2+1}.$$

If $X_1, \cdots, X_m$ is a sequence for which the process is terminated on the mth trial, the conditional probability that the interval of length l will cover ξ is clearly maximized by taking

$$(19) \quad Y = E(\xi \mid X_1, \cdots, X_m) = \frac{\sigma^2}{m\sigma^2+1} \sum_1^m X_i$$

and, by (18) this probability has the value $H(c_m)$ where H is defined by (6) and

$$(20) \quad c_m = \sqrt{m + \frac{1}{\sigma^2}} \frac{l}{2}.$$

Hence,

(21) $$\bar{\alpha}(\sigma, S) \leq \sum_{m=1}^{\infty} \bar{p}_m(\sigma, S) H(c_m).$$

From this and (10) we obtain

(22) $$\alpha(S) \leq \sum_{1}^{\infty} \bar{p}_m(\sigma, S) H(c_m).$$

This upper limit of $\alpha(S)$ and the lower limit of $n_0(S)$ given in (14) will be used later to prove that $S(\nu, c)$ is an optimum procedure.

4. Maximum value of $\sum_{1}^{\infty} \bar{p}_m(\sigma, S) H(c_m)$ subject to the condition that $\sum_{1}^{\infty} m\bar{p}_m(\sigma, S)$ does not exceed a given bound. We shall show that the maximum of $\sum_{1}^{\infty} \bar{p}_m(\sigma, S) H(c_m)$ subject to

$$E(n \mid \sigma, S) = \sum_{1}^{\infty} m\bar{p}_m(\sigma, S) \leqq \nu + a,$$

where ν is a positive integer and $0 \leq a < 1$, is obtained by choosing $\bar{p}_m(\sigma, S) = p_m^*$ defined by

(23) $$\begin{aligned} p_m^* &= 0 \text{ for } m < \nu \text{ or } m > \nu + 1 \\ p_\nu^* &= 1 - a \\ p_{\nu+1}^* &= a. \end{aligned}$$

For, suppose to the contrary that there exists a sequence $\{p_m\}$ such that the following conditions hold:

(24) $$p_m \geq 0, \qquad \sum_{1}^{\infty} p_m = 1$$

$$\sum_{1}^{\infty} m p_m \leq \nu + a = \sum_{1}^{\infty} m p_m^*$$

$$\sum_{1}^{\infty} p_m H(c_m) > \sum_{1}^{\infty} p_m^* H(c_m).$$

We have

(25) $$H(u) = \sqrt{\frac{2}{\pi}} \int_0^u e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_0^{u^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} dy.$$

Put

(26) $$C = H(c_{\nu+1}) - H(c_\nu) = \frac{1}{\sqrt{2\pi}} \int_{c_\nu^2}^{c_{\nu+1}^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} dy.$$

With the aid of $p_\nu = 1 - \sum_{m \neq \nu} p_m$, we obtain from the last two inequalities in (24)

$$(27) \qquad 0 < \sum_1^\infty (p_m - p_m^*)H(c_m) - C \sum_1^\infty (p_m - p_m^*)m = \sum_{m \neq \nu} (p_m - p_m^*)K_m$$

where

$$(28) \qquad K_m = H(c_m) - H(c_\nu) - (m - \nu)[H(c_{\nu+1}) - H(c_\nu)].$$

Clearly $K_{\nu+1} = 0$. Also, for $m < \nu$, since the integrand is a strictly decreasing function of y,

$$
(29) \qquad
\begin{aligned}
K_m &= (\nu - m) \int_{c_\nu^2}^{c_\nu^2+1} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \, dy - \int_{c_m^2}^{c_\nu^2} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \, dy \\
&< (\nu - m) \frac{l^2}{4} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \bigg|_{y=c_\nu^2} - (\nu - m) \frac{l^2}{4} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \bigg|_{y=c_\nu^2} = 0.
\end{aligned}
$$

Similarly for $m > \nu + 1$, $K_m < 0$. But $p_m^* = 0$ for $m \neq \nu$, $\nu + 1$ so that

$$(30) \qquad \sum_{m \neq \nu, 1 \, \nu+1} (p_m - p_m^*)K_m \leq 0$$

which contradicts (27) since $K_{\nu+1} = 0$.

Thus, we have shown that the inequality

$$(31) \qquad \bar{E}(n \mid \sigma, S) \leq \nu + a$$

implies the inequality

$$(32) \qquad \sum_1^\infty \bar{p}_m(\sigma, S)H(c_m) \leq (1 - a)H(c_\nu) + aH(c_{\nu+1}).$$

5. Proof that $S(\nu, c)$ is an optimum procedure. Since, according to (14) and (22)

$$(33) \qquad n_0(S) \geq \bar{E}(n \mid \sigma, S) \quad \text{and} \quad \alpha(S) \leq \sum_1^\infty \bar{p}_m(\sigma, S)H(c_m),$$

it follows from the result expressed in (31) and (32) that, for any procedure S satisfying the inequality

$$(34) \qquad n_0(S) \leq \nu + a,$$

we must have

$$(35) \qquad \alpha(S) \leq (1 - a)H(c_\nu) + aH(c_{\nu+1})$$

identically in σ. Since $H(u)$ is continuous, it follows that

$$(36) \qquad \alpha(S) \leq (1 - a)H\left(\sqrt{\nu} \, \frac{l}{2}\right) + aH\left(\sqrt{\nu + 1} \, \frac{l}{2}\right)$$

for any procedure S satisfying (34).

The right hand side of (36) is $\alpha[S(\nu, c)]$ where c is chosen so that

(37) $$1 - a = P\{\chi^2_{\nu-1} > c\}.$$

We use an indirect proof to show that $S(\nu, c)$ is an optimum procedure. Suppose to the contrary that there is a procedure S' such that

(38) $$\alpha(S') = \alpha[S(\nu, c)]$$

but

(39) $$n_0(S') < n_0[S(\nu, c)].$$

By (5) and (7), $\alpha[S(\nu, c)]$ is a continuous strictly increasing function of

$$\nu + 1 - P\{\chi^2_{\nu-1} > c\}$$

and this latter is $n_0[S(\nu, c)]$. If we choose ν', c' so that

(40) $$n_0(S') < \nu' + 1 - P\{\chi^2_{\nu-1} > c'\}$$
$$< \nu + 1 \quad -P\{\chi^2_{\nu-1} > c\},$$

it follows that

(41) $$\alpha[S(\nu', c')] < \alpha[S(\nu, c)] = \alpha(S').$$

But (41) and the first part of (40) contradict the result expressed in (34) and (36).

REFERENCES

[1] A. WALD, *Sequential Analysis*, John Wiley and Sons, 1947, section 11.2.
[2] CHARLES STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 243–258.
[3] G. B. DANTZIG, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ", *Annals of Math. Stat.*, Vol. 11 (1940), p. 186.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

A USEFUL CONVERGENCE THEOREM FOR PROBABILITY DISTRIBUTIONS

By Henry Scheffé

University of California at Los Angeles

In problems of establishing limiting distributions it is often apparent that the probability density $p_n(x)$ of a random variable X_n has a limit $p(x)$; throughout this paper $n = 1, 2, 3, \cdots$, and all limits are taken as $n \to \infty$. If $p(x)$ is the density of a random variable X, what we really care about then is whether the limits apply to probabilities, which involve integrals of the densities: Does $\lim Pr\{X_n \text{ in } S\} = Pr\{X \text{ in } S\}$ for all[1] Borel sets S, or, does

$$(1) \qquad \lim \int_S p_n(x) \, dx = \int_S p(x) \, dx \ ?$$

The question is thus one of taking a limit under an integral sign. Perhaps the most widely used justification of such a process is the following theorem of Lebesgue [1, p. 47; 2, p. 29]: If for a sequence $\{f_n(x)\}$ of integrable functions, $\lim f_n(x) = f(x)$ for almost all x in S, then a sufficient condition that

$$\lim \int_S f_n(x)dx = \int_S f(x) \, dx$$

is that there exist an integrable function $g(x)$ which uniformly dominates the sequence $\{f_n(x)\}$, that is, $|f_n(x)| \le g(x)$ for all n and all x in S, and $\int_S g(x) \, dx < \infty$. For example, in the excellent new treatise by Cramér the limiting form of the t-distribution is treated as follows [1, p. 252; other examples on pp. 369, 371]: For n degrees of freedom the t-variable has the density

$$(2) \qquad p_n(x) = c_n(1 + x^2/n)^{-\frac{1}{2}(n+1)},$$

where

$$(3) \qquad c_n = (n\pi)^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}(n + 1))/\Gamma(\tfrac{1}{2}n).$$

It is shown fairly easily that $\lim p_n(x) = p(x)$, the density of $N(0, 1)$, where

---

[1] In defining the convergence of a sequence of distributions to the distribution of a discontinuous random variable X it is desirable to modify this requirement so that it is demanded only of sets S which are continuity intervals of X [1, p. 83]. We are concerned here however only with the "absolutely continuous case" where X has a probability density $p(x)$.

$N(m, \sigma^2)$ denotes the normal distribution with mean m and variance σ^2. Then to prove

$$\lim \int_{-\infty}^{\xi} p_n(x) \, dx \; = \; \int_{-\infty}^{\xi} p(x) \, dx,$$

Cramér shows that $\{p_n(x)\}$ is uniformly dominated by an integrable function.

It is instructive to consider some examples where

(4) $$\lim \int_{-\infty}^{\xi} p_n(x) \, dx$$

does not equal

(5) $$\int_{-\infty}^{\xi} \lim p_n(x) \, dx.$$

In the examples (i), (ii), (iii), $\lim p_n(x) = 0$ for all x and hence (5) is zero for all ξ.

(i) $p_n(x) = 1$ for $-n - 1 < x < -n$, zero elsewhere. Then (4) equals 1 for all ξ.

(ii) $p_n(x) = 1/n$ for $-\frac{1}{2}n < x < \frac{1}{2}n$, zero elsewhere. Here (4) equals $\frac{1}{2}$ for all ξ.

(iii) $p_n(x) = 2n^2 x$ for $0 < x < 1/n$, zero elsewhere. Now (4) is zero for $\xi \leq 0$, unity for $\xi > 0$.

An example in which $\lim p_n(x) \neq 0$ is

(iv) $p_n(x) = \frac{1}{2}[h_n(x) + p_0(x)]$, where h_n is the p_n of one of the above examples and p_0 is a fixed density. Then $\lim p_n(x) = \frac{1}{2}p_0(x)$. Now (4) exceeds (5) by half the amount it did in the corresponding above example.

The essential features of these examples could be obtained with normal distributions but would involve a little more computation, for instance, $N(-n, 1)$, $N(0, n^2)$, $N(1/n, 1/n^4)$, for examples (i), (ii), (iii), respectively.

We note that in none of these examples is $\lim p_n(x)$ a density. This suggests that the trouble might perhaps be prevented by requiring that $\lim p_n(x)$ be a density—which happens in the case from which we started. This surmise is correct. We may formalize the situation as follows:

DEFINITION. *A function $f(x)$ will be called a density if it is non-negative and* $\int_R f(x) \, dx = 1$. *Here R denotes the whole space of x.*

The reader may think of a univariate density, where x is a real variable and R is the real axis, but theorem and proof run the same for a k-variate density, where x is a point in a k-dimensional Euclidean space R.

THEOREM[2]. *If for a sequence $\{p_n(x)\}$ of densities*

$$\lim p_n(x) \; = \; p(x)$$

---

[2] The hypotheses of this theorem, while perfectly adapted to applications in probability and statistics, would not seem the "natural" ones in real variable or measure theory. Professor A. P. Morse has remarked to the writer that, if the theorem has not been stated in this form before, it is at least an easy corollary of some more general results known in that field. Nevertheless our direct proof based only on the familiar Lebesgue theorem and using only

for almost all x in R, then a sufficient condition that

$$\lim \int_S p_n(x) \, dx = \int_S p(x) \, dx,$$

uniformly for all Borel sets S in R, is that $p(x)$ be a density.

PROOF. Let us write the difference

(6)
$$p_n(x) - p(x) = \delta_n(x).$$

Then

(7)
$$\delta_n(x) \to 0$$

for almost all x in R. Also

(8)
$$\int_S \delta_n \, dx = \int_S p_n \, dx - \int_S p \, dx,$$

and so it suffices to prove that $\int_S \delta_n \, dx \to 0$ uniformly for all S in R, where S

henceforth denotes a Borel set. If in (8) we let $S = R$ we get

(9)
$$\int_R \delta_n \, dx = 0$$

since p_n and p are densities. We now split the difference $\delta_n(x)$ into its positive and negative parts: Let

(10)
$$\delta_n^+ = \tfrac{1}{2}(\delta_n + |\delta_n|), \qquad \delta_n^- = \tfrac{1}{2}(\delta_n - |\delta_n|),$$

so that

$$\delta_n = \delta_n^+ + \delta_n^-, \qquad \delta_n^+ \geq 0, \qquad \delta_n^- \leq 0.$$

From (7) and (10), we find

(11)
$$\delta_n^- \to 0$$

for almost all x in R, and from (9),

(12)
$$\int_R \delta_n^+ \, dx + \int_R \delta_n^- \, dx = 0.$$

---

very simple manipulations may be of interest to readers of the *Annals*. Professor Morse also pointed out that the stronger result $\lim \int_S |p_n(x) - p(x)| \, dx = 0$ uniformly for all S, may be stated. This follows from our proof since

$$\int_S |p_n - p| \, dx = \int_S \delta_n^+ \, dx - \int_S \delta_n^- \, dx.$$

By virtue of (6), $\delta_n \geq -p$. Now if $\delta_n \leq 0, \delta_n^- = \delta_n \geq -p$, and if $\delta_n > 0, \delta_n^- = 0 \geq -p$, and hence in every case $0 \geq \delta_n^- \geq -p$. Since we now have $|\delta_n^-(x)| \leq p(x)$ and $\int_R p(x)\, dx = 1$, we may apply[3] the Lebesgue theorem to get

$$\lim \int_R \delta_n^-\, dx = \int_R \lim \delta_n^-\, dx.$$

The right member is zero because of (11). It then follows from (12) that $\lim \int_R \delta_n^+\, dx$ is also zero. The relations

$$0 \leq \int_S \delta_n^+\, dx \leq \int_R \delta_n^+\, dx \to 0,$$

$$0 \geq \int_S \delta_n^-\, dx \geq \int_R \delta_n^-\, dx \to 0$$

guarantee that the quantities $\int_S \delta_n^+\, dx$ and $\int_S \delta_n^-\, dx$ have the limit zero uniformly for all S, and hence the same is true of their sum (8).

Returning to the example (2), we remark that it is practically obvious that the second factor on the right has the limit $e^{-\frac{1}{2}x^2}$, but it is not quite so obvious that $\lim c_n = (2\pi)^{-\frac{1}{2}}$. This situation is typical of many applications where it is more difficult to evaluate the limit of "the" constant than the limit of the remaining factors, and one wonders after obtaining the latter limit whether the constant is not automatically forced toward the limit desired for it, and whether the direct calculation of its limit could not be avoided. Let us put the question as follows: Suppose that

$$\{p_n(x) = c_n f_n(x)\}$$

is a sequence of densities and that

$$p(x) = cf(x)$$

is also a density. Then if $\lim f_n(x) = f(x)$ for almost all x, may we conclude that $\lim c_n = c$? If so, we could then apply the above theorem without having evaluated the limit of the constant or produced a dominating function. Unfortunately the answer to this question is no, as shown by example (iv) above:

___

[3] Although our proof rests on the Lebesgue convergence theorem, this theorem is applied to $\delta(x)$ and not to $p_n(x)$. While in most cases of practical interest the sequence $\{p_n(x)\}$ is uniformly dominated by an integrable function, it is possible to devise a simple example where this is not true and yet our theorem applies: Let $p_n(x) = 1$ for $1/(n+1) \leq x \leq 1$ and for $a_n \leq x \leq a_{n+1}$, zero elsewhere, where $a_n = \sum_{i=1}^{n} 1/i$. Then sup $p_n(x) = 1$ for all $x > 0$, nevertheless $\lim p_n(x)$ is a density, namely that of the uniform distribution on $(0, 1)$.

If we let $f_n(x) = h_n(x) + p_0(x)$, and $f(x) = p_0(x)$, then $\lim f_n(x) = f(x)$, but $c_n = \frac{1}{2}$ and $c = 1$, hence $\lim c_n \neq c$. Employing the assumption that $p_n(x)$ and $p(x)$ are densities we see

$$1/c_n = \int_R f_n(x)\,dx, \qquad 1/c = \int_R f(x)\,dx,$$

and hence $\lim c_n = c$ if and only if

(13) $$\lim \int_R f_n(x)\,dx = \int_R \lim f_n(x)\,dx.$$

It follows that in such cases if we wish to establish a limiting distribution in the sense (1), we may either prove $\lim c_n = c$, or we may justify (13), say by producing a suitable dominating function, but we need not do both. No doubt the first alternative would be preferable at all but the most advanced levels of teaching or exposition.

REFERENCES

[1] H. Cramér, *The Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.
[2] S. Saks, *Theory of the Integral*, Stechert, New York, 1937.

AN EXPLICIT REPRESENTATION OF A STATIONARY GAUSSIAN PROCESS

By M. Kac[1] and A. J. F. Siegert

Cornell University and *Syracuse University*

1. In a paper which will soon appear in the *Journal of Applied Physics* [1] the authors have introduced methods of calculating certain probability distributions which are of importance in the theory of random noise in radio receivers.

The complexity of the physical problem and occasional uses of heuristic reasonings may have obscured some of the mathematical points. For this reason the authors felt that it may be worth while to illustrate one of the basic ideas on a simple but important example.

2. A stationary Gaussian process is a one parameter family $x(t)$ of random variables such that:

(a). $x(t)$ is normally distributed; the mean and the variance being independent of t

(b). the joint probability distribution of $x(t_1)$, $x(t_2)$, $\cdots$, $x(t_r)$ is multivariate Gaussian whose parameters depend only on the differences $t_j - t_k$.

[1] John Simon Guggenheim Memorial Fellow.

We assume, for the sake of simplicity, that the process is normalized, i.e.,

$$E\{x(t)\} = 0, \qquad E\{x^2(t)\} = 1$$

and we define the correlation function $\rho(\tau)$ by the usual formula

$$\rho(\tau) = E\{x(t)x(t + \tau)\}.$$

It is then well known[2] that a distribution function $\sigma(u)$ exists such that for all τ

(1)
$$\rho(\tau) = \int_{-\infty}^{\infty} \cos u\tau \, d\sigma(u).$$

3. Let $0 \le s, t \le T$ and consider the symmetric kernel

$$K(s, t) = \rho(s - t).$$

The fact that $\sigma(u)$ is non-decreasing implies that the kernel $\rho(s - t)$ is quasi-definite, i.e., for every L^2 function $g(t)$ on $(0, T)$ one has

$$\int_0^T \int_0^T g(s)\rho(s - t)g(t) \, ds \, dt \ge 0.$$

Thus the eigenvalues of the integral equation

(2)
$$\int_0^T \rho(s - t)f(t) \, dt = \lambda f(s)$$

are non-negative. Moreover, denoting by λ_j the eigenvalues and by $f_j(t)$ the corresponding normalized eigenfunctions of (2) we have by the classical theorem of Mercer (see [4], in particular part 6 of Ch. I) that

(3)
$$\rho(s - t) = \sum_j \lambda_j f_j(s)f_j(t),$$

where the series on the right is absolutely and uniformly convergent. It should be noted that in virtue of (1) $\rho(\tau)$ is a continuous function.

4. Let now $G_1, G_2, G_3, \cdots$ be independent, normally distributed random variables each having mean 0 and variance 1.

Consider the series

(4)
$$\sum_j \sqrt{\lambda_j} G_j f_j(t).$$

Since for each t we have

$$\sum_j (\sqrt{\lambda_j}f_j(t))^2 = \sum_j \lambda_j f_j^2(t) = \rho(0) = 1,$$

we infer that for each t the series (4) converges in the mean to a random variable $x(t)$. Moreover, by a theorem of Kolmogoroff [5], the series (4) converges, for each t, to $x(t)$ with probability 1.

---

[2] See [2]. The theorem in question (in a somewhat different form) seems to have been first established by N. Wiener in [3].

Thus we may write

(5)
$$x(t) = \sum_j \sqrt{\lambda_j} G_j f_j(t).$$

It is now easy to show that $x(t)$ thus defined is a stationary Gaussian process in $(0, T)$ with the correlation function $\rho(\tau)$.

In fact,

$$E\{x(s)x(t)\} = \sum_j \lambda_j f_j(s) f_j(t) = \rho(s - t), 0 \le s, t \le T,$$

and conditions (a) and (b) of section 2 follow from the well known properties of linear combinations of independent Gaussian random variables. Of course, we are dealing here with infinite linear combinations but the mean convergence noted above, is sufficient to justify the extension to our case.

5. It is more illuminating to think of the random variables G_j as measurable functions $G_j(\omega)$ defined on an abstract set Ω in which a Lebesgue measure has been established (the measure of the whole space being 1).

The representation (5) can then be written in the equivalent form

(6)
$$x(t, \omega) = \sum_j \sqrt{\lambda_j} G_j(\omega) f_j(t).$$

The equality, as established in section 4, holds for every t in the sense of mean convergence. Moreover, by the theorem of Kolmogoroff cited above, and by Fubini's theorem the equality (6) holds for almost every pair (t, ω), $(0 \le t \le T)$, in the sense of ordinary convergence.

Furthermore by Mercer's theorem (remember that $\lambda_j \ge 0$)

$$\sum_j \lambda_j = \int_0^T \rho(s - s) \, ds = T$$

and hence

$$\sum_j \lambda_j E\{G_j^2\} = \sum_j \lambda_j \int_\Omega G_j^2(\omega) = \sum_j \lambda_j = T < \infty.$$

Thus

$$\sum_j \lambda_j G_j^2(\omega)$$

converges for almost every ω and therefore the series

(7)
$$\sum_j \sqrt{\lambda_j} G_j(\omega) f_j(t)$$

converges in the mean for almost every ω.

Combining this fact with the observation that (7) converges almost everywhere to $x(t, \omega)$ we see that, for almost every ω, the series (7) converges in the mean to $x(t, \omega)$ and that consequently

(8)
$$\int_0^T x^2(t, \omega) \, dt = \sum_j \lambda_j G_j^2(\omega)$$

for almost every ω.

It should be noted that (8) could not, in general, be derived by just appealing to Parseval's relation. The main reason is that Parseval's relation holds only for complete orthonormal systems whereas the orthonormal system $\{f_n(t)\}$ of eigenfunctions may fail to be complete. If the kernel $\rho(s - t)$ is positive-definite (in which case all the eigenvalues are positive instead of just non-negative) then it is known that the eigenfunctions form a complete set. This actually, happens to be the case in most physical applications.

6. An important application of (8) is the calculation of the characteristic function of the distribution function of the random variable

(9)
$$I = \int_0^T x^2(t, \omega) \, dt.$$

In fact,

(10)
$$E\{\exp (i\xi I)\} = \prod_j E\{\exp (i\xi \lambda_j G_j^2)\} = \prod_j (1 - i\xi \lambda_j)^{-\frac{1}{2}}.$$

The probability density of I is the Fourier integral

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\xi I) \prod_j (1 - i\xi \lambda_j)^{-\frac{1}{2}} \, d\xi$$

which, unfortunately, in most cases cannot be calculated explicitely. If

$$\rho(\tau) = e^{-\beta|\tau|},$$

in which case the process is also Markoffian, the eigenvalues λ_j can be calculated explicitly[3] but in more complicated cases it is quite difficult to determine them.

7. If $\rho(\tau)$ is absolutely integrable and $\sigma(\mu)$ absolutely continuous then, setting

$$A(u) = \sigma'(u),$$

we have $A(u) \geq 0$ and

$$\rho(\tau) = \int_{-\infty}^{\infty} \cos u\tau A(u) \, du = \int_{-\infty}^{\infty} e^{iu\tau} B(u) \, du, \qquad B(u) \qquad A(u) + A(-u)$$

---

[3] See [6], in particular section 4. We take this opportunity to correct two misprints in this note. In the last formula on p. 64 M should be replaced by N. Also the limits of integration in formula (6) should be 0, s and s, $p + q$ instead of 0, $p + q$ and 0, $p + q$.

The N.D.R.C. Report 14-305 to which a reference is made has been declassified in the meantime. It contains results which originated both [1] and the present note.

[4] These and related results were stated in the abstract [7] by M. Kac. The paper is now being prepared for publication.

It can then be shown[4] that

$$\lim_{T \to \infty} \frac{1}{T} \sum_j \lambda_j^2 = 2\pi \int_{-\infty}^{\infty} B^2(u) \, du = \int_{-\infty}^{\infty} \rho^2(\tau) \, d\tau$$

and

$$\lim_{T \to \infty} \frac{1}{T} \sum_j \lambda_j^3 = (2\pi)^2 \int_{-\infty}^{\infty} B^3(u) \, du.$$

It follows now by standard methods that the characteristic function of

(11)
$$\frac{1}{\sqrt{T}} \left\{ \int_0^T x^2(t) \, dt - T \right\}$$

approaches, as $T \to \infty$,

$$\exp\left(-\frac{\sigma^2}{2} \xi^2 \right),$$

where

$$\sigma^2 = \int_{-\infty}^{\infty} \rho^2(\tau) \, d\tau.$$

Thus, as $T \to \infty$, the distribution of (11) becomes normal with mean 0 and variance σ^2.

REFERENCES

[1] M. Kac and A. J. F. Siegert, "On the theory of random noise in radio receivers with square law detectors." To appear in *Jour. of Applied Physics*.

[2] A. Khintchine, "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1934), pp. 604–615.

[3] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1930), pp. 117–258.

[4] G. Hamel, *Integralgleichungen*, Julius Springer, Berlin, 1937.

[5] A. Kolmogoroff, "Über die Summen durch den Zufall bestimmter unabhängiger Grössen," *Math. Ann.*, Vol. 99 (1928), pp. 309–319.

[6] M. Kac, "Random walk in the presence of absorbing barriers," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 62–67.

[7] M. Kac, "Distribution of eigenvalues of certain integral equations with an application to roots of Bessel functions," Abstract, *Bull. Amer. Math. Soc.*, Vol. 52 (1946), pp. 65–66.

APPROXIMATE FORMULAS FOR THE RADII OF CIRCLES WHICH INCLUDE A SPECIFIED FRACTION OF A NORMAL BIVARIATE DISTRIBUTION

By E. N. Oberg

University of Iowa

1. Introduction. Given the normal bivariate error distribution

(1)
$$\phi(x, y) = (1/2\pi\sigma_x \sigma_y)e^{-(x^2/2\sigma_x^2 + y^2/2\sigma_y^2)}.$$

The purpose of this paper is to present certain approximate formulas for the radii of circles whose centers are at the origin, which include a prescribed proportion, p, of errors. The formulas are, for given σ_x, σ_y, and p,

(2) $$R_1 = \sqrt{2\sigma_x \sigma_y \ln (1/[1 - p])}$$

(3) $$R_2 = \sqrt{(\sigma_x^2 + \sigma_y^2) \ln (1/[1 - p])}$$

and

(4) $$R_3 = (\sigma_x + \sigma_y)\sqrt{(1/2) \ln (1/[1 - p])}.$$

In section 3 we present tables of p', the true proportion of errors contained in circles whose radii are given by the above formulas. These tables reflect the goodness of approximation of each formula to the true radius, R, for $0.1 \leqq p \leqq 0.9$ and $0.5 \leqq \sigma_x/\sigma_y \leqq 0.9$. Also, a brief statement is included for the same range of p but with $0.1 \leqq \sigma_x/\sigma_y \leqq \cdot 4$.

2. The derivation of the formulas. The proportion p of errors that fall within an area A on the xy-plane is given by

(5) $$p = \int_A \varphi(x, y) \, dA.$$

If the area is bounded by any member of the family of elipses

$$x^2/\sigma_x^2 + y^2/\sigma_y^2 = \lambda^2,$$

the above integral may be evaluated directly. The result is

$$p = 1 - e^{-\lambda^2/2},$$

whence

$$\lambda^2 = 2\ln(1/[1 - p]).$$

Thus the ellipse with semi-axes

(6) $$\sigma_x\sqrt{2 \ln (1/[1 - p])}, \qquad \sigma_y\sqrt{2 \ln (1/[1 - p])},$$

measured from the origin along the x and y axes respectively, will include exactly the prescribed proportion of errors.

Frequently, however, it is desired to know which circles rather than which ellipses include a certain proportion of the errors. In this case it becomes difficult to obtain a formula for the true radius from (5) unless $\sigma_x = \sigma_y$ in which case R is given by either one of the formulas in (6). However, a natural approximation to make is to equate the area of a circle of radius, say R, to the area of the ellipse whose semi-axes are given in (6). This gives formula (2),

$$R_1 = \sqrt{2\sigma_x \sigma_y \ln (1/[1 - p])},$$

which can be expected to give a fairly close approximation to true R if σ_x is close to σ_y. If $\sigma_x \neq \sigma_y$, it has been shown that this formula underestimates true R which is undesirable in some applications [1]. That is, if R_1 is used to estimate, say the radius of a circle to include 50% of the errors ($p = .5$), it will give a value which includes less than the desired proportion. The first table in the last section gives a numerical verification of this fact.

To obtain formula (3) we consider formula (5) when A is a circle of radius R. We have

$$p = 4 \int_0^R \int_0^{\sqrt{R^2-x^2}} \varphi(x,\, y) \; dy \; dx.$$

By making the transformation $x = \sigma_x r \cos \theta$, $y = \sigma_y r \sin \theta$, and by carrying out the integration with respect to r the above formula becomes

$$p = 1 - (2/\pi) \int_0^{\frac{\pi}{2}} e^{-R^2/[(\sigma_x^2+\sigma_y^2)-(\sigma_y^2-\sigma_x^2)\cos 2\theta]} \; d\theta.$$

We let

$$\alpha = R^2/(\sigma_x^2 + \sigma_y^2), \qquad \beta = (\sigma_y^2 - \sigma_x^2)/(\sigma_y^2 + \sigma_x^2),$$

and

$$\sigma_x/\sigma_y = \epsilon \; ; \qquad \sigma_x < \sigma_y \, .$$

Then

$\alpha = R^2/\sigma_y^2(1 + \epsilon^2)$, and $\beta = (1 - \epsilon^2)/(1 + \epsilon^2)$, which is less than unity. This substitution will be helpful later in preparing tables. The fact that σ_x is taken less than σ_y places no limitation on the final results since we only have to interchange axes in the other case. The above integral may now be written as

$$p = 1 - (2/\pi) \int_0^{\pi/2} e^{-\alpha\beta/(1-\beta\cos 2\theta)} \; d\theta$$

(7)

$$= 1 - (2/\pi)e^{-\alpha} \int_0^{\pi/2} e^{-\alpha\beta\cos 2\theta/(1-\beta\cos 2\theta)} \; d\theta.$$

The integrand, say $F(\theta)$, in the last integral of (7) can be shown to be monotone increasing from $e^{-\alpha\beta/1-\beta}$ to $e^{\alpha\beta/1+\beta}$ as θ varies from 0 to $\pi/2$. Furthermore, it crosses the line $F(\theta) = 1$ somewhere in this interval and differs but little from it anywhere if the ratio σ_x/σ_y is close to 1, since β is then close to zero. If, therefore, we replace the integrand by $F(\theta) = 1$, we have $p = 1 - e^{-\alpha}$. Hence, if α is replaced by $R^2/(\sigma_x^2 + \sigma_y^2)$ and the result solved for R, we have formula (3),

$$R_2 = \sqrt{(\sigma_x^2 + \sigma_y^2) \ln (1/[1 - p])}.$$

Finally, formula (4),

$$R_3 = (\sigma_x + \sigma_y)\sqrt{(\tfrac{1}{2}) \ln (1/[1 - p])},$$

is obtained by taking the root-mean-square of the former two. This formula has certain advantages over the other two, the most obvious being that σ_x and σ_y enter linearly so that it is simple to evaluate for given σ_x, σ_y, and p. Secondly it will be seen by the tables and additional comments made in the last section that when $p = 0.5$,[1] R_3 overestimates true R by a slight amount for all

---

[1] This particular value of p gives the circular probable error. In this case $R_3 = 0.5887(\sigma_x + \sigma_y)$.

values of σ_x/σ_y, and it gives a fairly close approximation to true R for all p when $\sigma_x/\sigma_y \geqq 0.5$.

We close this section by making a few brief comments. In the first place, if any of the above formulas is to be computed from a sample of data, we take $\sqrt{\Sigma x^2/(n-1)}$ and $\sqrt{\Sigma y^2/(n-1)}$ as estimates of σ_x and σ_y respectively. Furthermore, we test the significance of these statistics by known formulas [2]. Finally, σ_x and σ_y may be replaced by $\sqrt{\dfrac{\pi}{2}}\, D_x$ and $\sqrt{\dfrac{\pi}{2}}\, D_y$, where D_x is the population mean deviation. Thus, for example,

$$R_3 = (D_x + D_y) \sqrt{\frac{\pi}{4} \ln (1/[1-p])}.$$

3. Tables. The first formula in (7) is useful in testing by means of numerical integration the goodness of approximation of the formulas R_1, R_2, and R_3 to

TABLE I

p' computed by means of formula R_1

σ_x/σ_y \\ p	.1	.2	.25	.3	.4	.5	.6	.7	.75	.8	.9
.5	.0988	.1951	.2425	.2893	.3815	.4720	.5615	.6508	.6960	.7422	.8408
.6	.0944	.1974	.2459	.2942	.3899	.4846	.5786	.6726	.7198	.7676	.8668
.7	.0997	.1987	.2480	.2972	.3950	.4924	.5894	.6864	.7350	.7838	.8835
.8	.0999	.1995	.2492	.2989	.3981	.4970	.5958	.6946	.7440	.7936	.8935
.9	.1000	.1999	.2498	.2997	.3996	.4993	.5991	.6988	.7483	.7986	.8985
1.0	.1000	.2000	.2500	.3000	.4000	.5000	.6000	.7000	.7500	.8000	.9000

the true value of R. We construct the tables by replacing R in α by one of these formulas, say formula R_1. This gives $\alpha = [2\epsilon/(1+\epsilon^2)][1/(1-p)]$. Since $\beta = (1-\epsilon^2)/(1+\epsilon^2)$, the right hand side of the formula in (7) may then be evaluated for a choice of ϵ and p giving a value we denote by p'. This is the actual proportion of errors that is included in the circle whose radius is R_1. If R_1 gave true R, then p' would be equal to p, so we may regard the difference of p and p' as a measure of the error arising when R_1 is used to estimate R.

In the following tables the chosen values of p and $\epsilon = \sigma_x/\sigma_y$ are listed in the first row and column respectively. The remainder of the tables include the corresponding values of p'.

We also have computed tables for $0.1 \leqq \sigma_x/\sigma_y \leqq 0.4$ which we have not included in this paper since for this range of values of σ_x/σ_y, all of the formulas give approximations that depart considerably from true R except R_3 when $p = 0.5$. For this case, $p' = .4776, .5004, .5109,$ and $.5120$ when $\sigma_x/\sigma_y = 0.1, 0.2, 0.3,$ and 0.4 respectively.

The difference between an entry in a column and the corresponding value of p at the head of the column reflects the error in estimating true R by means of R_1, R_2, and R_3. For example, if p is chosen as .5 and $\sigma_x/\sigma_y = .7$ then R_3 gives the radius of a circle which includes 50.13% of the errors. Thus R_3 overestimates true R by including .13% more of the errors.

By examining the tables it is seen that when $0.1 \leqq p \leqq 0.3$, R_1 gives the best approximation to the true value of R, while R_2 gives the poorest. If $0.4 \leqq p \leqq$

TABLE II
p' computed by means of formula R_2

σ_x/σ_y \ p	.1	.2	.25	.3	.4	.5	.6	.7	.75	.8	.9
.5	.1215	.2363	.2912	.3446	.4467	.5432	.6346	.7217	.7641	.8060	.8907
.6	.1116	.2202	.2732	.3255	.4274	.5261	.6218	.7146	.7600	.8050	.8949
.7	.1057	.2100	.2616	.3127	.4140	.5136	.6116	.7081	.7558	.8032	.8976
.8	.1022	.2039	.2546	.3051	.4056	.5055	.6048	.7034	.7525	.8014	.8991
.9	.1005	.2009	.2509	.3012	.4013	.5012	.6011	.7008	.7506	.8003	.8999
1.0	.1000	.2000	.2500	.3000	.4000	.5000	.6000	.7000	.7500	.8000	.9000

TABLE III
p' computed by means of formula R_3

σ_x/σ_y \ p	.1	.2	.25	.3	.4	.5	.6	.7	.75	.8	.9
.5	.1102	.2161	.2674	.3176	.4152	.5092	.6001	.6887	.7327	.7768	.8694
.6	.1056	.2089	.2597	.3100	.4090	.5059	.6009	.6944	.7408	.7872	.8817
.7	.1027	.2044	.2548	.3050	.4046	.5031	.6007	.6974	.7456	.7937	.8908
.8	.1011	.2017	.2519	.3020	.4018	.5013	.6003	.6991	.7483	.7976	.8963
.9	.1003	.2004	.2504	.3004	.4004	.5003	.6001	.6998	.7496	.7995	.8992
1.0	.1000	.2000	.2500	.3000	.4000	.5000	.6000	.7000	.7500	.8000	.9000

0.75, R_3 gives the best and R_2 the poorest; and if $0.8 \leqq p \leqq 0.9$ R_2 gives the best and R_1 the poorest. Thus formula R_3 for general use gives the best overall approximation. It may be remarked at this point that bounds for the true value of R can be found by applying two of the formulas, one of which overestimates while the other underestimates R. From the tables it is apparent that this can be done for values of $p \leqq 0.8$.

Finally, these formulas may be used to test roughly the normality of the data. For example, if proper estimates[2] of σ_x and σ_y are made from the data, and the

---

[2] See section 2.

corresponding value of R_3 computed for a chosen p, then approximately, the proportion p' of plotted errors should fall within the circle of radius R_3 .

REFERENCES

[1] HENRY SCHEFFÉ, *Armor and Ordinance Report No. A-224*, OSRD No. 1918, Div. 2, pp. 60–61.
[2] S. S. WILKS, *Mathematical Statistics*, Princeton Univ. Press, 1943, p. 131.

A NOTE ON THE EFFICIENCY OF THE WALD SEQUENTIAL TEST

BY EDWARD PAULSON

Institute of Statistics, University of North Carolina

The sequential likelihood ratio test of Wald for testing the hypothesis H_0 that the probability density function is $f(X, \theta_0)$ against the one-sided alternative H_1 that the function is $f(X, \theta_1)$ has been shown [1] to have the optimum property of minimizing the expected number of observations at the two points $\theta = \theta_0$ and $\theta = \theta_1$. Tables showing the actual magnitude of the percentage saving of this sequential procedure compared with the classical "best" non-sequential test have been calculated (see [1], page 147) for the normal case when

$$f(X, \theta) = \frac{1}{\sqrt{2\pi}} \exp \frac{-(X - \theta)^2}{2} .$$

In this note we will show that when θ_1 is close to θ_0 , the percentage saving is independent of the particular function $f(X, \theta)$ and the particular values θ_1 and θ_0, so that the tables mentioned above can be used to show the percentage saving for any one-sided sequential test involving a single parameter, provided $f(X, \theta)$ satisfies some weak restrictions.

Let $f(X, \theta)$ be the probability density function of a random variable. Let $E_i(n)$ denote the expected value (when $\theta = \theta_i$) of the number of independent observations required by the Wald sequential procedure to test the hypothesis H_0 that $\theta = \theta_0$ against $\theta = \theta_1 = \theta_0 + \Delta$ with probabilities α of rejecting H_0 when $\theta = \theta_0$ and β of accepting H_0 when $\theta = \theta_1$. Let N be the number of independent observations required to achieve the same probabilities α and β for testing the hypothesis $\theta = \theta_0$ against $\theta = \theta_1$ by the most powerful non-sequential test. Let U_α and U_β be defined by the relations

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{U_\alpha}^{\infty} \exp\left\{-\frac{t^2}{2}\right\} dt$$

and

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{U_\beta}^{\infty} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

We will prove the following theorem:

$$\operatorname*{Limit}_{\Delta=\theta_1-\theta_0\to 0}\left\{\frac{E_0(n)}{N}\right\} = -2\,\frac{\left\{\alpha\log\left(\dfrac{1-\beta}{\alpha}\right) + (1-\alpha)\log\left(\dfrac{\beta}{1-\alpha}\right)\right\}}{(U_\alpha + U_\beta)^2}$$

provided $f(X, \theta)$ *satisfies the following conditions:*

(A) $\displaystyle\int_{-\infty}^{\infty} f(X, \theta)\, dx$ *can be differentiated twice under the integral sign with respect to* θ.

(B) *All four of the integrals*

$$\int_{-\infty}^{\infty}\left\{\frac{f''(x, \theta^*)}{f(x, \theta^*)} - \left[\frac{f'(x, \theta^*)}{f(x, \theta^*)}\right]^2\right\} f(x, \theta_0)\, dx,$$

$$\int_{-\infty}^{\infty}\frac{f'(x, \theta_0)}{f(x, \theta_0)} f'(x, \theta^*)\, dx,$$

$$\int_{-\infty}^{\infty}\left[\frac{f'(x, \theta_0)}{f(x, \theta_0)}\right]^2 f(x, \theta^*)\, dx,$$

$$\int_{-\infty}^{\infty}\frac{f'(x, \theta_0)}{f(x, \theta_0)} f(x, \theta^*)\, dx,$$

are continuous functions of θ^* *at* $\theta^* = \theta_0$. *A sufficient condition for* (B) *is that all the integrals be uniformly convergent with respect to* θ^* *in some interval* $\theta_0 \le \theta^* \le \theta_0 + \Delta$, *and all the integrands be continuous functions of* X *and* θ^*. *A similar theorem holds regarding the limit of* $\left\{\dfrac{E_1(n)}{N}\right\}_{\Delta\to 0}$.

The proof is as follows: From [1], we know that

$$E_0(n) = \frac{\alpha\log\left(\dfrac{1-\beta}{\alpha}\right) + (1-\alpha)\log\left(\dfrac{\beta}{1-\alpha}\right)}{E_0(z)} + o(1),$$

where

$$z = \log\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]$$

and $o(1) \to 0$ as $\Delta \to 0$.

Now

$$E_0(z) = \int_{-\infty}^{\infty}\left[\log\left(\frac{f(x, \theta_1)}{f(x, \theta_0)}\right)\right] f(x, \theta_0)\, dx,$$

$$= \int_{-\infty}^{\infty}[\log f(x, \theta_0 + \Delta)] f(x, \theta_0)\, dx - \int_{-\infty}^{\infty}[\log f(x, \theta_0)] f(x, \theta_0)\, dx.$$

Expanding $\log f(x, \theta_0 + \Delta)$ in a Taylor series about $\Delta = 0$, we have

$$\log f(x, \theta_0 + \Delta) = \log f(x, \theta_0) + \Delta \frac{f'(x, \theta_0)}{f(x, \theta_0)} + \frac{\Delta^2}{2}\left[\frac{ff'' - f'^2}{f^2}\right]_{\theta=\theta_0} + \frac{\Delta^2}{2} R_1 ,$$

where

$$\theta_0 \leq \theta^* \leq \theta_0 + \Delta, \qquad f' = \frac{\partial f(x, \theta)}{\partial \theta} , \qquad f'' = \frac{\partial^2 f(x, \theta)}{\partial \theta^2} ,$$

and

$$R_1 = \left[\frac{ff'' - f'^2}{f^2}\right]_{\theta=\theta_0}^{\theta=\theta^*}.$$

From assumption (A) we find that

$$\int_{-\infty}^{\infty} f'(x, \theta_0) \, dx = 0 \text{ and } \int_{-\infty}^{\infty} f''(x, \theta_0) \, dx = 0,$$

while from assumption (B)

$$\int_{-\infty}^{\infty} R_1 f(x, \theta_0) \, dx \to 0 \text{ as } \Delta \to 0.$$

Therefore

$$E_0(z) = -\frac{\Delta^2}{2}\left[\int_{-\infty}^{\infty}\left[f\left(\frac{f'}{f}\right)^2\right]_{\theta=\theta_0} dx + o\,(1)\right].$$

To find N for the most powerful non-sequential test, we make use of the fact (see [2]) that an asymptotically most powerful test for one-sided alternatives is given by a region of the type

$$U_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{i=N} \frac{f'(x_i, \theta_0)}{f(x_i, \theta_0)} \geq K.$$

When $\Delta \to 0$, $N \to \infty$, and since U_N is the sum of N independent variates with a finite second moment, the distribution of $\dfrac{U_N - E(U_N)}{\sigma_{U_N}}$ approaches that of a normal variate with zero mean and unit variance. Hence we find the N required for a test with Type I and Type II errors α and β by solving for N from the relations

(1)
$$\frac{K}{\sqrt{E_0\left(\dfrac{f'}{f}\right)^2_{\theta=\theta_0}}} = U_\alpha$$

and

(2)
$$\frac{K - \sqrt{N}\, E_1\left(\dfrac{f'}{f}\right)_{\theta=\theta_0}}{\sqrt{E_1\left(\dfrac{f'}{f}\right)^2_{\theta=\theta_0} - \left[E_1\left(\dfrac{f'}{f}\right)_{\theta=\theta_0}\right]^2}} = -U_\beta$$

Now let $y = \left(\dfrac{f'}{f}\right)_{\theta = \theta_0}$, and we find from (1) and (2) that

$$N = \left[\frac{U_\alpha \sqrt{E_0(y^2)} + U_\beta \sqrt{E_1(y^2) - [E_1(y)]^2}}{E_1(y)}\right]^2.$$

Now

$$E_1(y) = \int_{-\infty}^{\infty} \frac{f'(x, \theta_0)}{f(x, \theta_0)} f(x, \theta_1) \, dx$$

$$= \Delta \int_{-\infty}^{\infty} \frac{f'(x, \theta_0)}{f(x, \theta_0)} f'(x, \theta_0) \, dx + \Delta \int_{-\infty}^{\infty} \frac{f'(x, \theta_0)}{f(x, \theta_0)} [f'(x, \theta)]_{\theta=\theta_0}^{\theta=\theta^*} \, dx$$

$$= \Delta E_0 y^2 [1 + o(1)] \text{ from assumption } B.$$

Proceeding in a similar manner, we find

$$[U_\alpha \sqrt{E_0(y)^2} + U_\beta \sqrt{E_1(y)^2 - [E_1(y)]^2}]^2 = E_0(y^2)[U_\alpha + U_\beta(1 + o(1))]^2.$$

We now have

$$\frac{E_0(n)}{N} = \frac{\Delta^2 [E_0(y^2)]^2 (1 + o(1))^2}{E_0(y^2)[U_\alpha + U_\beta(1 + o(1))]^2} \times \frac{\alpha \log\left(\dfrac{1 - \beta}{\alpha}\right) + (1 - \alpha) \log\left(\dfrac{\beta}{1 - \alpha}\right)}{-\dfrac{\Delta^2}{2}[E_0(y^2) + o(1)]}$$

therefore

$$\lim_{\Delta \to 0} \left\{\frac{E_0(n)}{N}\right\} = -2 \frac{\left[\alpha \log\left(\dfrac{1 - \beta}{\alpha}\right) + (1 - \alpha) \log\left(\dfrac{\beta}{1 - \alpha}\right)\right]}{(U_\alpha + U_\beta)^2}.$$

REFERENCES

[1] A. WALD, "Sequential tests of statistical hypotheses", *Annals of Math. Stat.*, Vol. 16 (1945).

[2] A. WALD, "Some examples of asymptotically most powerful tests", *Annals of Math. Stat.*, Vol. 12 (1941).

A NOTE ON THE POISSON-CHARLIER[1]
FUNCTIONS

By C. TRUESDELL

Naval Ordnance Laboratory

The polynomials $p_n(m, z)$ given by the definition

$$(1) \qquad p_n(m, z) \equiv (-)^m e^z z^{-m} \frac{d^n}{dz^n} [e^{-z} z^m],$$

---

called the Poisson-Charlier polynomials, and the associated function $\psi_n(m, z)$ given by the definition

$$(2) \qquad \psi_n(m, z) \equiv p_n(m, z)\psi_0(m, z),$$

$$(3) \qquad \psi_0(m, z) \equiv \frac{e^{-z} z^m}{m!},$$

occur in statistics. Doetsch [1] has devoted a memoir to them, and they are noticed in Szegö's *Orthogonal Polynomials* (pp. 33–34).

I suggest that they are most directly and easily studied in connection with the "F-equation"

$$(4) \qquad \frac{\partial}{\partial z} F(z, \alpha) = F(z, \alpha + 1),$$

whose properties and application to various special functions I have summarized in a recent note [2]. Using the theorems of that note, which I shall cite by number, I shall now generalize the Poisson-Charlier polynomials and sketch the speediest derivation of their most interesting formal properties.

Greek letters shall represent unrestricted real numbers, while Latin letters shall represent integers.

From the existence theorem for the F-equation (Theorem 4) we know that there exists an integral function of z, $F_\beta(z, \alpha)$, which satisfies the F-equation and the condition

$$(5) \qquad F_\beta(0, \alpha) = \cos(\alpha + \beta)\pi \binom{\beta}{-\alpha}.$$

From the uniqueness theorem for the F-equation (Theorem 4) it follows that

$$(6) \qquad F_\beta(z, n - \beta + \tfrac{1}{2}) = 0,$$

$$(7) \qquad F_\beta(z, n) = 0, \qquad n > 0.$$

From the general power series solution for the F-equation (Theorem 4) we have the formula

$$(8) \qquad F_\beta(z, \alpha) = \cos(\alpha + \beta)\pi \binom{\beta}{-\alpha} {}_1F_1(\alpha; \beta + \alpha + 1; z).$$

We now define the Poisson-Charlier functions in general by the formulas

$$(9) \qquad p_\beta(\alpha, z) \equiv \Gamma(\alpha + 1)z^{-\alpha}F_\beta(z, -\alpha),$$

$$(10) \qquad \psi_\beta(\alpha, z) \equiv \frac{e^{-z} z^\alpha}{\Gamma(\alpha + 1)} p_\beta(\alpha, z).$$

From the formulas (6) and (7) we see that [1, p. 263]

$$(11) \qquad \psi_\beta(-n, z) = 0, \qquad n > 0;$$

(12) $\psi_\beta(-n + \beta - \frac{1}{2}, z) = 0$, $p_\beta(-n + \beta - \frac{1}{2}, z) = 0$.

From the formula (8) we see that

(13) $p_\beta(\alpha, z) = \cos(\beta - \alpha)\pi \dfrac{\Gamma(\beta + 1)}{\Gamma(\beta - \alpha + 1)} z^{-\alpha} {}_1F_1(-\alpha; \beta - \alpha + 1; z)$,

whence it follows at once that

(14) $p_\beta(m, z) = \cos \beta\pi \displaystyle\sum_{k=0}^{m} \binom{m}{k}\binom{\beta}{k} k!(-z)^{-k}$.

This is the usual explicit expression for the Charlier polynomials [1, p. 257]. From formula (13) we see that

(15) $p_0(-\alpha, z) = \dfrac{\sin 2\alpha\pi}{2\pi}\,\Gamma(1 - \alpha)z^\alpha\gamma(\alpha, z)$.

In the indeterminate case when α is a negative integer we see from the formula (14) that

(16) $p_0(m, z) = 1$, $m \geqq 0$.

Hence

(17) $\psi_0(-\alpha, z) = \dfrac{\sin 2\alpha\pi}{2\pi}\,e^{-z}\gamma(\alpha, z)$,

(18) $\psi_0(m, z) = \dfrac{e^{-z} z^m}{m!}$.

From the definition (10) we now see that

(19) $\psi_\beta(m, z) = p_\beta(m, z)\psi_0(m, z)$,

a generalization of the formula (2). From the formula (13) and the definition (10) we see that

(20) $\psi_\alpha(\beta, z) = \cos(\beta - \alpha)\pi \dfrac{\Gamma(\alpha + 1)}{\Gamma(\beta + 1)\Gamma(\alpha - \beta + 1)} e^{-z} {}_1F_1(-\beta; \alpha - \beta + 1; z)$.

Then by Kummer's first transformation,

(21) $\psi_\alpha(\beta, z) = \cos(\beta - \alpha)\pi \dfrac{\Gamma(\alpha + 1)}{\Gamma(\beta + 1)\Gamma(\alpha - \beta + 1)} {}_1F_1(\alpha + 1; \alpha - \beta + 1; -z)$,

from which it follows from the power series formula for solutions of the F-equation (Theorem 4) that $\psi_\alpha(\beta, z)$ is a solution of the F-equation (4).

We now have two different solutions of the F-equation based on the Poisson-Charlier functions:

(A) $F(z, \alpha) = e^z\psi_\beta(-\alpha, z)$.

(B) $$F(z, \alpha) = \psi_\alpha(\beta, z).$$

From the F-equation it is evident that

(22) $$\psi_n(\beta, z) = \frac{\partial^n}{\partial z^n} \psi_0(\beta, z),$$

whence we at once deduce the formula (1). Applying Taylor's theorem for the F-equation (Theorem 8) to the solution (B) we see that [1, p. 259]

(23) $$\psi_\alpha(\beta, z + h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \psi_{\alpha+n}(\beta, z);$$

putting α equal to zero we find that

(24) $$-\frac{\sin 2\beta\pi}{2\pi} e^{-z-h} \gamma(-\beta, z + h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \psi_n(\beta, z),$$

and, more specially [1, p. 260]

(25) $$\left(1 + \frac{h}{z}\right)^m e^{-h} = \sum_{n=0}^{\infty} \frac{h^n}{n!} p_n(m, z).$$

Applying the same theorem to the solution (A) we obtain the formula

(26) $$e^h \psi_\beta(\alpha, z + h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} \psi_\beta(\alpha - n, z),$$

whence we recover the formula (11) by putting α equal to zero.

Applying Theorem 9 to the solution (B) yields the result

(27) $$\sum_{n=0}^{\infty} t^n \psi_{\alpha+n}(\beta, z) = \int_0^\infty e^{-\theta} \psi_\alpha(\beta, z + \theta t) \, d\theta,$$

which contains as a special case the formula

(28) $$\sum_{n=0}^{\infty} t^n p_n(m, z) = (1 + t)^{-m-1} \left(\frac{t}{z}\right)^m e^{z(1+1/t)} \left[m! - \gamma\left(m + 1, z\left(1 + \frac{1}{t}\right)\right)\right].$$

Appell's generating expansion (see Theorem 10, part C or [3, p. 120]) applied to the solution (A) yields the result

(29) $$\sum_{n=0}^{\infty} \psi_\beta(n, z + y) t^n = e^{z(t-1)} \sum_{n=0}^{\infty} \psi_\beta(n, y) t^n;$$

hence

(30) $$\sum_{n=0}^{\infty} \frac{t^n}{n!} p_\beta(n, z + y) = e^{(zt)/(z+y)} \sum_{n=0}^{\infty} \left(\frac{yt}{z + y}\right)^n \frac{p_\beta(n, y)}{n!}.$$

Putting y equal to zero and using the formula (13) we see that

(31) $$\sum_{n=0}^{\infty} \frac{t^n}{n!} p^\beta(n, z) = e^t \left(1 - \frac{t}{z}\right)^\beta \cos \beta\pi.$$

Comparing this result with the formula (25) we see that

(32) $$(-)^n p_n(m, z) = (-)^m p_m(n, z).$$

It would be possible to proceed in this same fashion and discover many other formal properties of the Poisson-Charlier functions, but it is perhaps easier to notice from the formula (13) that

(33) $$p_\beta(\alpha, z) = \cos (\beta - \alpha)\pi\Gamma(\alpha + 1)z^{-\alpha}L_\alpha^{(\beta-\alpha)}(z).$$

$L_\delta^\gamma(x)$ being Laguerre's function suitably generalized for complex lower index [4, p. 53]. By means of this formula every relationship involving Laguerre functions may be translated into one involving Poisson-Charlier functions.

REFERENCES

[1] G. Doetsch, "Die in der Statistik seltener Ereignisse auftretenden Charliersche Polynome, und eine damit zusammenhangende Differentialdifferenzengleichung", *Math. Ann.*, Vol. 109 (1934), pp. 257–266.

[2] C. Truesdell, "On the functional equation $\frac{\partial}{\partial z} F(z, \alpha) = F(z, \alpha + 1)$", *Proc. Nat. Acad. Sci. U. S. A.*, April, 1947.

[3] P. Appell, "Sur une classe de polynomes", *Ann. Ecole Norm.*, Vol. 9 (1880), pp. 119–144.

[4] E. Pinney, "Laguerre functions in the mathematical foundations of the electromagnetic theory of the paraboloidal reflector", *Jour. Math. Phys. M. I. T.*, Vol. 25 (1946), pp. 49–79.

ABSTRACTS OF PAPERS

1. Random Variables with Comparable Peakedness. Z. W. BIRNBAUM, University of Washington.

Let U and V be random variables with symmetrical distributions, i.e. with $P(U \leqq -T) = P(U \geqq T)$ and $P(V \leqq -T) = P(V \geqq T)$ for all $T \geqq 0$. The random variable U shall be called more peaked than V if $P(|U| \geqq T) \leqq P(|V| \geqq T)$ for all $T \geqq 0$. Let X_1, Y_1 and X_2, Y_2 be two pairs of independent random variables such that X_i is more peaked than Y_i for $i = 1, 2$. Then under certain additional conditions $X = X_1 + X_2$ is more peaked than $Y = Y_1 + Y_2$.

2. On Optimum Tests of Composite Hypotheses with One Constraint. ERICH L. LEHMANN, University of California, Berkeley.

The problem studied is that of finding all similar and bisimilar test regions of composite hypotheses, and of obtaining the most powerful of these regions. Various results are obtained for distributions which admit sufficient statistics with respect to their parameters. Applications are made to the hypothesis specifying the value of the circular correlation coefficient in a normal population, and certain hypotheses concerning scale and location parameters in exponential and rectangular populations.

3. Estimation of a Distribution Function by Confidence Limits. FRANK J. MASSEY, JR., University of California, Berkeley.

Let $x_1, x_2, \cdots, x_n$ be the results of n independent observations, having the same cumulative distribution function $F(x)$. Form the function $S_n(x) = k/n$ where k is the number of observations less than or equal to x. A confidence band $S_n(x) \pm \lambda/\sqrt{n}$ will be used to estimate $F(x)$. To determine the confidence coefficient it is necessary to find $Pr\{\max \sqrt{n} \mid S_n(x) - F(x) \mid \leqq \lambda/\sqrt{n}\}$. It is sufficient to consider x uniformly distributed in the interval $(0, 1)$. Let $\lambda\sqrt{n} = s/t$ where s and t are integers. Then $S_n(x)$, to stay in the band $F(x) \pm \lambda/\sqrt{n}$, can only pass through certain lattice points above $x = i/tn$, $i = 1, 2, \cdots, tn$. The probability of $S_n(x)$ passing through a particular sequence of these points is given by the multinomial law, and this can be summed over all permissible sequences. Limiting distributions have been given by A. Kolmogoroff, and by N. Smirnoff. It is desired to test the hypothesis $F(x) = F_0(x)$ against alternatives $F(x) = F_1(x)$. Using the criterion: reject $F_0(x)$ if

$$\max_x \sqrt{n} \mid F_0(x) - S_n(x) \mid > \lambda$$

the probability of first kind of error can be controlled by choice of λ. A lower bound to the probability of second kind of error against alternatives such that $\max \sqrt{n} \mid F_0(x) - F_1(x) \mid \geqq \Delta$ is given. This lower bound approaches one as $n \to \infty$. Thus the test is consistent.

4. A Note on Sequential Confidence Sets. CHARLES STEIN, Columbia University.

This paper generalizes a paper of Stein and Wald, appearing in the *Annals of Math. Stat.*, Sept., 1947.

Let $\{X_i\}$, $(i = 1, 2, \cdots)$, be a sequence of random variables whose distribution depends on an unknown parameter θ. Sequential confidence sets are determined by a rule indicating

when to stop sampling and a rule giving the confidence set as a function of the sample. It is desired that, for each sample point, the confidence set should be one of a specified class S, that the probability of covering the true parameter should be $\geq \alpha$, and that the least upper bound of the expected number of observations should be minimized. If X_i are independent with the rectangular distribution on $(0, \theta)$ and S consists of all intervals of the form $(\theta_0, k\theta_0)$ with k fixed and θ_0 a function of the sample, the optimum sequential procedure is the classical non-sequential procedure. If the X_i are independently and identically distributed in accordance with a multivariate normal distribution with known covariance matrix Σ but unknown mean θ, and the confidence sets are to be of the form $(\theta - \theta_0)'\Sigma^{-1}(\theta - \theta_0) = r$, r fixed, θ_0 a variable p-dimensional vector, a similar result holds, provided the desired confidence coefficient α is not excessively small.

5. Explicit Solution of the Problem of Fitting a Straight Line when Both Variables are Subject to Error for the Case of Unequal Weights. ELIZABETH L. SCOTT, University of California, Berkeley.

Let α, β and $\xi_i, (i = 1, 2, \cdots, s)$, be unknown fixed numbers and let $\eta_i = \alpha + \beta\xi_i$. For each value of i there exist m_i measurements x_{ij} of ξ_i and n_i measurements y_{ik} of η_i, $(j = 1, 2, \cdots, m_i ; k = 1, 2, \cdots, n_i)$. The variables x_{ij} and y_{ik} are normally distributed about ξ_i and η_i with variances σ_1^2/u_i and σ_2^2/v_i respectively, where the weights u_i and v_i are known but σ_1^2 and σ_2^2 are unknown. The numbers m_i and n_i are bounded (usually small) while s increases indefinitely. Thus $\alpha, \beta, \sigma_1^2$ and σ_2^2 appear as structural parameters and the ξ_i as incidental parameters. (See paper by J. Neyman and E. L. Scott to appear in *Econometrica*.) Modified maximum likelihood equations (MMLE) yielding consistent estimates of the structural parameters are tedious to solve when the products $m_i u_i$ and $n_i v_i$ depend on i. The main result of this paper consists in proving that the varying $m_i u_i$ and/or $n_i v_i$ can be treated as constants. Let w_1 and w_2 be the harmonic means of $m_i u_i$ and $n_i v_i$, respectively. Now, MMLE's written with $m_i u_i = w_1$ and $n_i v_i = w_2$ yield consistent estimates of α and β. The asymptotic variances are also found. An application is made to certain problems of astronomy.

6. Unbiased Estimates with Minimum Variance. CHARLES STEIN, Columbia University.

Let X be a random variable distributed in the space R according to one of the p.d.f.'s $\varphi(x \mid \theta)$, where θ is an unknown parameter, and let $g(\theta)$ be a real-valued function of θ. Let $B(\theta)$ be the set of all x such that $\varphi(x \mid \theta) > 0$ but $\varphi(x \mid \theta_0) = 0$, and S the set of all θ such that $B(\theta)$ has probability 0 when θ is the true parameter value. Let

$$\psi(x \mid \theta) = \varphi(x \mid \theta_0)/\varphi(x \mid \theta_0 \text{ and } A(\theta_1, \theta_2) = E\{\psi(X \mid \theta_1) \psi(X \mid \theta_2) \mid \theta_0\}$$

for θ_1, θ_2 in S. Suppose $A(\theta_1, \theta_2)$ is everywhere finite and there exists a set function λ of bounded variation over S such that $\int_S A(\theta_1, \theta_2) d\lambda(\theta_1) = g(\theta_2)$. Then an estimate of $g(\theta)$, unbiased for all θ in S and having minimum variance at θ_0 is given by $f(x) = \int_S \varphi(x \mid \theta) d\lambda(\theta)/\varphi(x \mid \theta_0)$. The minimum variance is $\int_S g(\theta) d\lambda(\theta) - [g(\theta_0)]^2$. If the definition of $f(x)$ is modified at a set having probability 0 when $\theta = \theta_0$, the properties on S and at θ_0 remain unchanged. Under mild restrictions this alteration can be carried out so as to make $f(x)$ an unbiased estimate of θ for all S. The results are related to the work of Fisher, Dugué, Rao, and Bhattacharyya on the amount of information.

7. Sufficient Statistics and a System of Partial Differential Equations. (A Contribution to the Neyman-Pearson Theory of Testing Hypotheses.) Pre-

liminary Report. ERICH L. LEHMANN, University of California, Berkeley, AND HENRY SCHEFFÉ, University of California, Los Angeles.

In the Neyman-Pearson theory of testing hypotheses the problem of the existence and determination of similar regions has been treated under two approaches: (1) Assuming the existence of a set of sufficient statistics for the nuisance parameters; (2) Assuming that the probability density satisfies a certain system of partial differential equations. By solving the differential equations it is now shown that they imply the existence of sufficient statistics for the nuisance parameters. Knowledge of the form of the solution of the differential equations permits simplification of the known theory of optimum tests (type B, B_1, etc.) as well as some generalization.

8. Power Function of the Analysis of Variance and Covariance Test of a Normal Bivariate Population. W. M. CHEN, University of California, Berkeley.

The problem of finding the power function of the analysis of variance and covariance test of a normal bivariate population, $\rho = 0$ and $\sigma_1 = \sigma_2$, by means of principle of likelihood was reduced to the determination of the distribution function $P(L)$ of the following moment problem:

$$\int_0^1 L^k \, d\, P(L) = \frac{(1-a)^{(n-1)/2}}{\Gamma\left(\dfrac{n-1}{2}\right)} \sum_{r=0}^{\infty} \frac{a^r}{r!} \Gamma\left(\frac{n-1}{2} + r\right) M_{k,r}, \qquad (k = 1, 2, \cdots),$$

where

$$M_{k,r} = \frac{4^k \, \Gamma(n-1+r) \, \Gamma\left(\dfrac{n-2+2k}{2}\right) \Gamma\left(\dfrac{n-1+2k}{2} + r\right)}{\Gamma\left(\dfrac{n-2}{2}\right) \Gamma\left(\dfrac{n-1}{2} + r\right) \Gamma\left(n-1+2k+r\right)},$$

and a, the argument of the power function, lies in the interval $(0, 1)$ and vanishes only when the hypothesis tested is true. The moment problem was found and solved by rather tricky methods. The result is

$$P(L) = \frac{(1-b)^{(n-1)/2}}{\Gamma\left(\dfrac{n-1}{2}\right)} \sum_{s=0}^{\infty} \frac{b^s}{s!} \Gamma\left(\frac{n-1}{2} + s\right) I_L\left(\frac{n-2}{2}, s+1\right)$$

where $b = \left(\dfrac{a}{1-a}\right)^2$.

9. A Mathematical Model of the Relation between White and Yolk Weights of Birds' Eggs. G. A. BAKER, University of California, Davis.

The purpose of such a model is to find a rational method of estimating a "best line" in some sense which will represent the relation between white and yolk weights for some or all species of birds. From data at hand it appears that birds within species may differ in means and variances of weights and that the yolk and white weights are positively correlated. Yolk and white weights within a species are functions of egg number. The standard deviations of yolk and white weights for different species are approximately proportional to mean values. The "true" means for yolk and white weights for different species do not lie on a line because of biological differences between species with the same egg size. The standard deviation of species deviations from a straight line depend on the size of the egg (may be proportional to a weighted sum of the yolk and white weights). If sampling

variances are sufficiently small they may be neglected and a straight line fitted assuming both variables subject to error and non-uniform variance. The practicality of maximum likelihood estimates is considered.

10. Statistical Analysis for a New Procedure in Sensitivity Experiments. A. M. MOOD, Iowa State College, AND W. J. DIXON, University of Oregon.

In the language of biological assay the sensitivity experiment investigates the proportion of subjects that respond to a given concentration, x, of a certain chemical. It is assumed that only one test may be made on each subject. The new procedure is characterized by a change in x for each successive test, depending on the result of the preceding test. x is reduced to the next lower of a fixed set of concentrations for the next test if there is no response and is increased to the next higher concentration if there is a response. Observations are thus concentrated near the mean and few tests are made for values of x where a very large or very small proportion of subjects would respond. Assuming x is normally distributed, approximate maximum likelihood estimates are obtained for the mean and standard deviation of x. These assume a form which is simple to compute. Choice of optimum increments of x for various situations is investigated.

11. The Relation of Inbreeding to Calf Mortality. W. M. REGAN, S. W. MEAD, AND P. W. GREGORY, University of California, Davis.

An analysis of calf mortality in the University of California dairy cattle breeding experiment is presented. Calves up to 4 months of age that were born singly are included in the study. Only those stillbirths and abortions from cows free from Brucellosis and health and reproductive abnormalities were considered. A total of 774 Jersey and 258 Holstein calves were included. Calves were classified according to inbreeding coefficients as follows Class I, the controls 0 0 to 0.1249; Class II, 0.125 to 0.2448; Class III, 0.245 to 0.3749; and Class IV, 0.375 and over. There was no relation between the number of abortions and the degree of inbreeding. The stillbirths, too few to be statistically significant, tend to increase as the coefficient of inbreeding increased. Following birth, however, mortality was correlated with inbreeding of both males and females but for the males it was greater than for the females in Classes III and IV, but the difference is hardly significant. The Jerseys tended to be less viable than the Holsteins. Some of the increased mortality of the more highly inbred animals could be accounted for by the action of two lethal genes; one controlling an anomaly of the liver, the other an anomaly of the heart; there was no plausible explanation for most of it. Within sex, inbreeding class, and breed there was considerable variation in the mortality of the progeny of different sires. Some of these differences were statistically significant.

12. Observations on Designs for Cooperative Field Tests. P. A. MINGES, University of California, Davis.

In California conditions vary so greatly between the principal production areas that it is necessary to establish experimental plots in each of the areas if reliable information is to be obtained regarding cultural practices. Most of these tests must be conducted on ranches in cooperation with growers and local agricultural extension agents. The designs of these tests should be relatively simple, the arrangement should be adjustable to work into the growers' cultural practices and to permit the obtaining of yield records with a minimum of interference to the growers' operations, yet the design must be adequate to yield valid data. The randomized block design has proved the most useful, although paired plots, factorials, split-plots and Latin squares have been used successfully under certain conditions. The Latin square design is useful when a two-way variation is expected, other-

wise it is not usually very efficient. Where yield data are of prime importance, for ureplica-tions have been considered most practical. In tests such as variety trials when factors other than yields are important, two replications may be adequate. The size of the plot has been varied to fit the crop, conditions of the field, and known soil variability. Plots two rows wide and 50 to 135 feet long often have been used, frequently without guards between plots. Since it is desirable to include checks (untreated controls) in most tests, small plots will reduce the loss to the growers when the treatments prove beneficial. The information derived from these tests is of most interest to growers and county agents so the data should be presented in tables that are easily read. The variability figure which is confusing to most people probably can best be presented as the least significant difference.

13. Population Genetics. N. H. HOROWITZ, California Institute of Technology.

Population genetics attempts to describe the effects on the genetical structure of Mende-lian populations of factors such as mutation, selection, migration, and random fluctuations due to sampling errors. These diverse elements are brought under a common viewpoint by considering their effects on gene frequencies. Since change in gene frequency is the ele-mentary process of evolution, the above factors are causal agents of evolution. Mathe-matical models illustrating the interplay of the various elements have been constructed by Wright, Haldane, and Fisher. The nature of Mendelian inheritance is such that gene frequencies remain constant in large populations not subject to net mutation, selection, or migration pressures. Unbalanced pressures initiate evolutionary changes which continue until equilibrium is reached at a new level of gene frequencies. Equilibrium frequencies are determined by opposing pressures—e.g., opposing mutation rates, mutation opposed by selection, etc. Equilibrium, stable or unstable, is also possible under selection alone. In small populations, sampling errors among the gametes produce random fluctuations in gene frequencies which, superimposed on the equilibrium values, result in probable distributions of frequencies. The latter provide a mechanism for the evolution of characters, especially biochemical syntheses, which depend on the simultaneous action of a number of individually non-adaptive genes.

14. The Choice of Inspection Stringency in Acceptance Sampling by Attributes. J. L. HODGES, JR., University of California, Berkeley.

In acceptance sampling by attributes, the probability p that an item will be defective is taken to be a function $g(x, y)$ of the quality x of the population and the stringency y of inspection. Let n, the number of items inspected, be fixed, and reject if the number of defectives is $\geq k$. It may then be possible to satisfy a condition on the power function with different values of k, by adjusting y properly. This paper is concerned with the choice of k and y in such situations. A criterion is given, and it is shown that the criterion is approximately satisfied by $k = [ng(x_0, \hat{y})]$ where x_0 separates acceptable and non-acceptable values of x, and $\hat{y}$ maximizes

$$\frac{\partial g(x_0, y)}{\partial x} \bigg/ \sqrt{g(x_0, y)[1 - g(x_0, y)]}.$$

An asymptotic property of this approximation is shown. The method is applied to two examples: (a) testing the mean bacterial density x of a liquid by the dilution method, y being the volume of liquid incubated, and (b) testing the variance x of a normally dis-tributed dimension of known mean m by applying gauges set at $m \pm \dfrac{1}{y}$. The approximate solution is found to be satisfactory in both cases for $m = 20$.

15. The Application of Learning Curves to Industrial Planning. Preliminary Report. JAMES R. CRAWFORD, Lockheed Aircraft Corporation.

Learning curves are significant factors of analysis in industries producing quantities of less than 20,000 units of a given article. Ship-building and airframe manufacture are the two largest industries in this class. Learning curves occur where job costs are kept either by individual unit or by lot, and also where achievement is measured against a standard. Cost per unit plots against ordinal unit number as a straight line on logarithmic graph-paper. Learning curves are used to supplement time-studies, determine the capacity of tooling, layout of budgets, and for estimating and bidding. The experience of individual workers and management are reflected in these analyses. The slope of the learning curve is related to the amount to be learned. Plateaus occur which are related to the hiring of new workers and to the relaxing of control measures. Other consistent minor patterns occur which are related to specific conditions. Equations have been derived and tables computed for five related forms of the learning curve. Graphic methods are satisfactory except for bidding. This study covers a simple approach to an important problem of industrial management. The findings in the industrial field may benefit research in the field of the psychology of learning.

16. Relative Effects of Inbreeding and Selection in Poultry. W. O. WILSON, University of California, Davis.

Egg production rate, fertility, hatchability, and chick mortality records from the Iowa State College Poultry Department's inbreeding project were studied. Statistics which were calculated from the data included simple and partial regression of traits on inbreeding, estimates of heritability by correlation between paternal half-sibs and by daughter-dam regressions, and selection differentials. The net genetic gain or loss in merit per generation was considered to be the sum of the product of selection differentials and heritability, plus the product of regression of trait on inbreeding and increase in amount of inbreeding. The amount of inbreeding that can be done in each of the traits was estimated when there was no net loss or gain. Of the traits studied, the rank was in the following order: Hatchability, chick mortality, fertility, and egg production.

17. The Rate of Genetic Gain in Egg Production in Progeny-Tested Flocks as a Function of the Interval between Generations. EVERETT R. DEMPSTER AND I. MICHAEL LERNER, University of California, Berkeley.

The rate of genetic gain in a character for which selection is practiced depends in addition to the intensity of selection on (1) the accuracy of selection, and (2) the average interval between generations. These factors are not independent and exercise a pull in opposite directions. Through the application of Wright's technique of path coefficients comparisons can be made between the expected rates of genetic gain in populations containing varying proportions of breeding animals of different ages. The methods used involve the estimation of correlations between genotypes, and various selection indexes based on individual, sib and progeny records in inculled populations as well as in populations whose range has been restricted by previous selection. From these estimates the relative efficiencies of different age distribution schemes of a breeding population can be determined. A specific solution for such a situation in a flock bred for egg production will be presented as an illustration of the problems and methods used in the study of the genetics of populations under artificial selection.

18. Statistical Criteria of the Effectiveness of Selective Procedures. Preliminary Report. R. F. JARRETT, University of California, Berkeley.

The "validity coefficient," the standard error of estimate, the index of predictive efficiency, the "selection ratio" of Taylor and Russell, Johnson's Gamma, and other statistical devices have been suggested as indices of the effectiveness of selective programs. These devices all suffer from the deficiency that they do not permit a satisfactorily precise estimate of the dollar value of the increased output expected from the selection program and thus leave unsettled the question as to whether or not the cost of such a program is justified. The relationship between the correlation coefficient on the one hand and the mean value of Y for an unselected population (Y being an objective output-type criterion), the standard deviation of Y for an unselected population, and the mean value of Y for the upper Np individuals selected on the basis of their high performance on the selective test X on the other hand, provides the basis for estimating the increase in the mean output of a group of workers selected on the basis of a testing program yielding any specified validity coefficient with the criterion Y. Increase in productivity of selected workers is shown to be a function of the validity coefficient, the rigorousness of selection, and the coefficient of variability of the output criterion among "unselected" employees.

19. Approaches to Univocal Factor Scores. Preliminary Report. J. P. GUILFORD, University of Southern California.

In spite of the fact that univocal factor scores are badly needed for various reasons, it appears to be impossible by present methods to construct pure tests for some common factors. Recourse must therefore be made to statistical control of component variances. It is desirable to derive each factor score from a minimum number of tests. The availability of a few univocal tests makes this requirement fairly easy to satisfy. Such tests serve well as suppression variables for their common-factor variances where not wanted in other tests. Several principles may be invoked as objectives: (1) to maximize the desired variance in the impure test, (2) to reduce the undesired variance to zero, or (3) to minimize the undesired variance without intolerable loss of the desired variance. A secondary objective is to assure a combining weight of $+1.00$ for the test measuring the desired factor. Equations for achieving the objectives have been derived and the limitations and implications of each procedure have been noted. By means of statistical control, the situation seems hopeful for the achievement of univocal scores for a fairly large number of unique psychological variables. There are implications for experimental psychology as well as for vocational testing.

20. A Note on the Problem of Binary Stars. ELIZABETH L. SCOTT, University of California, Berkeley,

This paper concerns some of the problems of Trumpler (see next abstract). ξ_{ij} is the radial velocity of the i-th star, $i = 1, 2, \cdots, s$, at t_j selected at random, $j = 1, 2, \cdots, n$. x_{ij}, measurement of ξ_{ij}, is $N(\xi_{ij}, \sigma_i)$. ξ_{ij} is random with distribution $c(k_i^2 - (\xi_{ij} - \xi_{i0})^2)^{-\frac{1}{2}}$ where $k_i \geqq 0$ and ξ_{i0} are unknown. (1) Test of hypothesis that $k_i = 0$. Case (i) σ_i known. Whatever the exact test T, its power $\beta_T(k)$ has derivative $\beta'_T(0) = 0$. Test maximizing $\beta''_T(0)$ is that of Trumpler with criterion $S^2 = \sum_{j=1}^{n} (x_{ij} - x_{i\cdot})^2 > \chi^2 \sigma_i^2$. Case (ii) σ_i unknown. Whatever the exact test τ, $\beta_\tau^{(m)}(0) = 0$, $m = 1, 2, 3$. Test maximizing $\beta_\tau^{(4)}(0)$ is Trumpler's test $\sum_{j=1}^{n} (x_{ij} - x_{i\cdot})^4 > \left[\sum_{j=1}^{n} (x_{ij} - x_{i\cdot})^2 \right]^2 C$. (2) Let $\sum_{j=1}^{n} (\xi_{ij} - \xi_{i0})^2 = 2\lambda_i \sigma^2$. For constant velocity stars $\lambda = 0$. For others it is a random variable. Since, given $\lambda = 0$, S^2 is distributed as non-central χ^2, an integral equation connects the distributions of S^2 and λ. Its solution yields an estimate of the proportion of constant velocity stars. After estimating the distribution of λ, the level of significance can be estimated and also the

number n of measurements so that the proportion of constant velocity stars declared variable will be less than p, specified in advance.

21. Statistical Problems of Spectroscopic Binaries. ROBERT J. TRUMPLER, University of California, Berkeley.

Spectroscopic Binaries are stars whose radial velocities, as measured by the Doppler shift of spectral lines, show a periodic variation. The first problem is to obtain a statistical criterion for deciding whether a star with several radial velocity measures, made at different times, has a high probability (larger than a specified limit) of variable velocity and should be announced as an object worthy of further study. The second problem is to find the percentage of variable velocity stars among a large list of stars with several radial velocity measures for each star. From the distribution of standard errors only the percentage of cases where the velocity variation exceeds a certain limit can be ascertained. The third problem is concerned with those stars for which a binary orbit has been determined. The statistical distribution of these binary systems according to mean distance between the two stars and the ratio of their masses can be evaluated within certain limits.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Kenneth J. Arrow has been appointed Research Associate of the Cowles Commission.

Dr. W. D. Baten, formerly of Michigan State College, is now Chief, Operations Branch, Planning Section, Air Defense Command, Mitchel Field, New York.

Dr. Paul T. Bruyere is now Chief of the Medical Records and Statistics Branch, Army Institute of Pathology, Office of the Surgeon General, War Department.

Dr. A. C. Cohen received his discharge from the Army, with the rank of Lieutenant Colonel, at the beginning of the spring quarter, and returned to his former position at Michigan State College. He has accepted a position at the University of Georgia beginning with the 1947 summer session there.

Dr. Hallett H. Germond has resigned from his position as professor of mathematics at the University of Florida. He is now Director of Research for the S. W. Marshall firm of Consulting Engineers, in New York City.

Dr. Meyer A. Girshick, formerly with the Department of Agriculture, is now with the Douglas Aircraft Company in Santa Monica, California.

Dr. Clyde H. Graves has accepted a position as Operations Analyst, Operations Analysis, Air Defense Command, Mitchell Field, New York.

Dr. E. J. Gumbel has been appointed to an Associate Professorship at Brooklyn College.

Dr. Trygve Haavelmo has returned to Norway, and is at the University Institute of Economics, Oslo.

Mr. Joseph O. Harrison, Jr., is now employed as a Mathematician in the Computing Branch of the Ballistic Research Laboratories, Aberdeen Proving Ground.

Dr. Wassily Hoeffding has accepted a psoition as Research Associate, The Institute of Statistics, University of North Carolina, Chapel Hill.

Mr. Cyrus A. Martin is now an administrative analyst and statistician, assisting Chief of Personnel Control of Signal Corps, in Washington, D. C.

Mr. Jack I. Northam has accepted an Assistant Professorship in the Department of Mathematics, Kansas State College, Manhattan, beginning with the 1947 summer session.

Professor Henry Scheffé, who has been on leave for the past year, returned to his position in the Engineering Department, University of California at Los Angeles, in June.

Mr. Edward M. Schrock has accepted a position as Quality Control Engineer with the General Electric Company at their Erie Works, Erie, Pa.

Mr. Jerome R. Steen, who has been manager of Quality Control Engineering

with the Sylvania Electric Products in Emporium, Pa., has now transferred with the same company to Flushing, New York.

Professor Emeritus Irving Fisher, of Yale University, died April 29, 1947, at the age of eighty.

In connection with the Atlantic City meeting of the American Chemical Society, April 14-18, 1947, a symposium on *Statistical Methods in Experimental and Industrial Chemistry* was held, in which several members of the Institute of Mathematical Statistics took part. The following program was presented Tuesday morning and afternoon, April 15:

(1) *Introductory Remarks.* B. L. Clarke.
(2) *The Management Viewpoint* George Smith.
(3) *A New Technique for Testing the Accuracy of Analytical Data.* W. J. Youden.
   Discussion: Grant Wernimont, R. F. Moran, John Mandel, and Roland H. Noel
(4) *Design of Experiments in Industrial Research.* Hugh M. Smallwood.
(5) *Statistical Training for Industry.* Samuel S. Wilks.
   Discussion: John Tukey, E. V. Lewis, Churchill Eisenhart, and C. West Churchman.

Preliminary Actuarial Examinations

Prize Awards

The winners of the prize awards offered by the Actuarial Society of America and the American Institute of Actuaries to the nine undergraduates ranking highest in the combined score on Part 1 and Part 2 of the 1947 Preliminary Actuarial Examinations are as follows:

First Prize of $200

James H. Chung	*University of Toronto*

Additional Prizes of $100

James F. A Biggs	*Yale University*
George Y. Cherlin	*Rutgers University*
Frank H David	*Harvard University*
Thomas M Galt	*University of Manitoba*
Charles F Pinzka	*Rutgers University*
Philip C. Rapp	*University of Buffalo*
Morton K Schwartz	*Brown University*
James G C. Templeton	*University of Toronto*

The two actuarial organizations have authorized a similar set of nine prize awards for the 1948 Examinations.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1. *Language Aptitude Examination*
(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning).

Part 2. *General Mathematics Examination*
(Algebra, trigonometry, coordinate geometry, differential and integral calculus).

Part 3. *Special Mathematics Examination*
(Finite differences, probability and statistics).

The 1948 Examinations will be administered by the College Entrance Examination Board at centers throughout the United States and Canada on May 14–15, 1948.

Correction

In the Directory of Members published in Vol. XVII, No. 4 (December 1946) Professor Joseph Kampé de Feriet's name is listed in the F's under Feriet. It should have appeared in the K's, under Kampé de Feriet.

New Members

The following persons have been elected to membership in the Institute (March 1 to May 30, 1947):

Adams, Joe K. Ph. M. (Wisconsin) Graduate student and half-time instructor in Psychology, Graduate College, Princeton University, Princeton, N. J.

Adams, Walter B. Communications Analyst, Civil Aeronautics Admin., Dept. of Commerce, *8253 S. Ingleside Ave., Chicago 19, Ill.*

Aitken, Alexander C. D.Sc. (Edinburgh) Professor of Mathematics, University of Edinburgh, *23 Stirling Road, Edinburgh 5, Scotland*

Brambilla, Francesco Ph.D. (Univ. L. Bocconi) Lecturer in Math. Statistics, Institute of Statistics, Università L. Bocconi, 6 via Panzacchi, Milano, Italy

Brown, George Middleton, D.Sc. (Michigan) Asst. Prof. of Math., Mich. State College, East Lansing, Mich., *633 Cherry Lane*

Bueno, Luiz de Freitas, E.E. (Mackenzie Coll.) Professor da Universidade de São Paulo, Brazil, *Rua Itambé 341, Casa 13*

Burke, Cletus J., M.A. (U.C.L.A.) Res. Ass't, Univ. of Iowa, Iowa City, Iowa, *118 Riverside Park*

Cameron, Joseph M., M.S. (N. Car. State) Room 302 South Building, National Bureau of Standards, Washington, D. C.

Carpenter, Osmer M.S. (Iowa State) Instructor, Mathematics Department, Iowa State College, Ames, Iowa

Castellani, Maria D.Sc. (Rome) Visiting Professor, Department of Mathematics, University of Kansas City, Kansas City 4, Mo

Chernoff, Herman Sc.M. (Brown) National Research Council Pre-Doctoral Fellow, *3003 Wallace Ave., Bronx 67, N. Y.*

Clark, Stanley M.Ed. (Saskatchewan) Student and teaching assistant, *1301-7th St., S.E., Minneapolis 14, Minn.*

Cover, John H. Ph.D. (Columbia) Director, Bureau of Business and Economic Res., Univ. of Maryland, College Park, Md.

Dailey, John T. M.S. (N. Texas Teachers Coll.) Res. Psychologist (Aviation), Psychological Res. and Examining Unit, Sqn. E, Indoctrination Div., Air Training Command, San Antonio, Texas

Darling, Donald A. Ph.D. (Calif. Inst. Tech.) Teaching Ass't, Calif. Inst. of Technology, Pasadena 4, Calif. (As of July 1947, Dept. of Math., Cornell Univ., Ithaca, N. Y.)

Darmois, Georges D.Sc. (Paris) Prof. à la Faculté des Sciences de Paris, *7 Rue de l'Odéon, Paris 6, France*

Davies, J. Alfred M.A. (Alabama) Statistician, Design Eng. Section, General Electric Co., *708 Hill Avenue, Ownesboro, Kentucky*

Dunnett, Charles W. M.A. (Toronto) Student, 1044 John Jay Hall, Columbia Univ., New York 27, N. Y.

Egermayer, Frantisek Sc.D. (Charles Univ., Prague) Chief of Section, State Statistical Office, *2 Bělského, Prague VII, Czechoslovakia.*

Fickenscher, Edgar H. A.B. (Calif.) Graduate student and teaching ass't, Univ. of Calif., *1490 Acton St., Berkeley 2, Calif.*

Fraga, Constantino G. Jr. (Sao Paulo) Head, Dept. of Statistics, Instituto Agronomico, Campinas (S.P.), Brazil.

Frank, Elmore J. B.A. (Chicago) Instr. in Statistics, Ill. Institute of Tech., and Statistician, Commercial Res. Dept., Armour and Co., *5423 Maryland Ave., Chicago 15, Ill.*

Frisch, Ragnar Ph.D. (Oslo) Professor, University Institute of Economics, Oslo, Norway.

· **Geary, Robert C.** D.Sc. Superintending Officer, Statistics Branch, Dept. of Industry and and Commerce, 27 Leeson Park, Dublin, Ireland.

Goodman, John R. M.S. (Iowa State) Head, Sampling Section, Survey Res. Center, Univ. of Mich., Ann Arbor, Mich.

Gutman, Pierre M.A. (Columbia) Student, *7 Mountain Ave., Maplewood, N. J.*

Hartline, H. K. M.D. (Johns Hopkins) Assoc. Prof. of Biophysics, Johnson Res. Foundation, Univ. of Pennsylvania, 36th and Spruce Sts., Philadelphia, Pa.

Hartog, Jacob A. (Rotterdam) Rockefeller Fellow, *25 Follen St., Cambridge, Mass.*

Jacobs, Marcus A.B. (Penn.) Health Statistician, *4439 S. 36th St., Arlington, Va.*

Jeeves, Terry A. A.B. (Calif.) Teaching ass't in math., Univ. of Calif., *2511 Hearst Ave., Berkeley 9, Calif.*

Kempthorne, Oscar M.A. (Cambridge, England) Res. Assoc. Prof., Statistical Lab., Iowa State College, Ames, Iowa

Kendall, David G. M.A. (Oxford) Fellow, Magdalen Coll., Oxford, England

Kent, Leonard M.B.A. (Chicago) Instr. in Statistics, School of Business, Univ. of Chicago, Chicago 37, Ill.

Kupperman, Morton B.S. (C.C.N.Y.) Statistician, Office of the Surgeon General, War Dept., *2829-27th St., N.W., Washington 8, D. C.*

Lhati, Elizabeth L. M.A. (Michigan) Statistician, Bur. of Measurement and Guidance, Carnegie Institute of Technology, Pittsburgh 13, Pa.

Levine, Harry D. B.S. (Chicago) Instr., Long Island Univ., *164 W. 96 St., New York 25, N. Y.*

Lichtenstein, Morris B.A. (Michigan) Statistician, *4811 N. Capitol St., N.E., Washington 11, D. C.*

McMillan, Brockway Ph.D. (Mass. Inst. Tech.) Member, Technical Staff, Bell Telephone Labs., Murray Hill, N. J.

Marshall, Andrew W. Student, *5757 University Ave., Chicago 37, Ill.*

Metzner, Charles A. Ph.D. (Wisconsin) Study Director, Survey Research Center, Univ. of Michigan, Ann Arbor, Mich.

Norton, John W. B.S. (California) Lab. Supervisor, Union Oil Co. of Calif., *5529 Macdonald Ave., Richmond, Calif.*

Otter, Richard Ph.D. (Indiana) Instructor, Fine Hall, Princeton Univ., Princeton, N. J.

Passos, Helena Rocha Penteado Diretor de Divisão do Depto. Estadual de Estatística de São Paulo, *Avenida Angélica, 160, Apto. 6, São Paulo, Brazil*

Priest, Edward I. B.S. (Columbia) Student in mathematics, *1204 E. 55th St., Chicago, Ill.*

Quensel, Carl-Erik Fil.Dr. (Lund) Prof. at the University, Lund, Sweden, *Linnegatan 14*

Rankin, Mozelle M.A. (Ohio State) Ass't Instructor, Ohio State Univ., *107-14th Ave., Columbus 1, Ohio*

Robb, Richard A. D.Sc. (Glasgow) Mathematics Lecturer and Mitchell Lecturer in Statistics, Univ. of Glasglow, Glasgow, W. 2, Scotland.

Ruist, Erik Fil.kand. (Stockholm) Amanuens, Industriens utredningsinstitut, Stockholm 16, Sweden

Shani, Inder M. M.A. (Punjab) Rothamsted Experimental Station, Harpenden, Herts, England.

Schneider, B. Aubrey Sc.D. (Johns Hopkins) Ass't Director, Dept. of Statistics and Special Services, American Cancer Society, 47 Beaver St., New York 4, N. Y.

Seitz, Jiri Ph.D. (Prague) Kouřimská 8, ČSR, Praha XII, Czechoslovankia.

Simaika, Jacques B. Ph.D. (London) Lecturer, Faculty of Science, Fuad I University, Abbassia, Cairo, Egypt.

Slatin, Benjamin M.A. (Columbia) Jr. Analyst, Econometric Institute, *179 Peshine Ave., Newark 8, N. J.*

Suydam, Bergen R. A.B. (N.Y. State Coll. for Teachers) Graduate student, Columbia University, *1 W. 706 St., Shanks Village, Orangeburg, N. Y.*

Tashmuhamed, Sarymsakov Ph.D. (Moscow) President of the Academy of Sciences of Uzb.SSR, Professor of the University, Tashkend, *ul. Abdulli Tukaeva I, Tashkent, USSR*

Travers, Robert M. W. Ph.D. (Columbia) Examiner, and Assoc. Prof. of Education, Bureau of Psychological Services, Univ. of Mich., Ann Arbor, Mich.

Weiner, Sidney B.S. (C.C.N.Y.) Student, New York University Graduate School, *1539 East 17th St., Brooklyn 30, N. Y.*

Wezelman, Sol M. A.B. (Michigan) Graduate student, University of Michigan, Ann Arbor, Mich., *2432 Burt St., Omaha, Nebr.*

Wishart, John D.Sc. (London) Reader in Statistics, School of Agriculture, Cambridge, England

REPORT ON THE NEW YORK MEETING OF THE INSTITUTE

The Twenty-Sixth Meeting of the Institute of Mathematical Statistics was held in New York City on Thursday, April 24, and Friday, April 25, 1947, and was co-sponsored by the American Mathematical Society. This meeting was devoted to a program on *Stochastic Processes and Noise*. The attendance of 190 persons included the following 75 members of the Institute:

F. A. Acton, C. B. Allendoerfer, F. L. Alt, T. W. Anderson, Jr., L. A. Aroian, W. D. Baten, Robert Bechhofer, J. H. Bigelow, D. H. Blackwell, Paul Boschan, G. W. Brown, R. S. Burington, B. H. Camp, E. W. Cannon, A. G. Carlton, K. L. Chung, P. C. Clifford, D. D. Cody, Harald Cramér, H. B. Curry, J. H. Curtiss, R. L. Dietzold, J. L. Doob, Jacques Dutka, Churchill Eisenhart, Benjamin Epstein, Will Feller, M. M. Flood, Bernard Friedman, C. P. Gerschenson, H. H. Goode, C. H. Graves, E. J. Gumbel, T. E. Harris, Millard Hastay, L. H. Herbach, P. G. Hoel, Mark Kac, R. D. Keeney, T. C. Koopmans, William Kruskal, Jack Laderman, J. E. Lieberman, S. B. Littauer, Melitta Lowy, P. J. McCarthy, Brockway Mc-Millan, Frederick Mosteller, L. F. Nanni, P. M. Neurath, G. E. Noether, M. L. Norden, C. O. Oakley, P. S. Olmstead, G. B. Price, J. S. Rhodes, John Riordan, Selby Robinson, Frank Saidel, Arthur Sard, F. E. Satterthwaite, G. R. Seth, C. E. Shannon, Jack Sherman, W. A. Shewhart, Rosedith Sitgreaves, Andrew Sobczyk, Milton Sobel, Emma Spaney, C. M. Stein, J. W. Tukey, D. F. Votaw, Jr., B. T. Weber, S. S. Wilks, Jacob Wolfowitz.

The first session, was held on Thursday morning, with Professor Carl Allendoerfer of Haverford College serving as chairman. The following program was presented:

Stochastic Processes—

Description, Professor J. L. Doob, Columbia University
Estimation, Professor Will Feller, Cornell University
Prediction, Professor N. Wiener, Massachusetts Institute of Technology

This meeting was concluded with a discussion by Dr. H. W. Bode, Bell Telephone Laboratories, Professor Mark Kac, Cornell University, and Professor A. Wald, Columbia University.

Dr. S. O. Rice, Bell Telephone Laboratories, was chairman of the Thursday afternoon session. The following program was presented:

Stochastic Processes in Some Applications—

In Economics, Dr. T. Koopmans, Cowles Commission
In Insurance, Professor H. Cramér, Yale University
In Cosmic Radiation, Professor N. Arley, Princeton University
In Nuclear Physics, Dr. S. M. Ulam, Los Alamos Laboratory

The final session was held on Friday morning with Professor J. W. Tukey of Princeton University as chairman. The program was as follows:

Different Ways of Describing Noise—

By a Noise Spectrum, Dr. C. E. Shannon, Bell Telephone Laboratories
By a Single Function, Mr. J. E. Bigelow, Institute for Advanced Study
By Many Functions, Professor Mark Kac, Cornell University
Round Table on Interrelations, Messrs. Shannon, Bieglow, Kac, and Rice

<div align="right">

P. S. DWYER,
Secretary.

</div>

REPORT ON THE APRIL MEETING OF THE INSTITUTE IN ATLANTIC CITY

The Twenty-Seventh Meeting of the Institute of Mathematical Statistics was held in cooperation with the Eastern Psychological Association, on Saturday morning, April 26, 1947, in Atlantic City. This meeting was a *Round Table on Certain Recent Statistical Developments*, and its attendance of approximately 100 persons included the following 9 members of the Institute:

F. S. Acton, J. W. Dunlap, Benjamin Epstein, Irving Lorge, P. J. McCarthy, Frederick Mosteller, P. J. Rulon, F. E. Satterthwaite, and Emma Spaney.

Professor Bernard Riess of Hunter College was chairman of the meeting. The following program was presented:

Papers: *Sequential Analysis.*
Dr. Irving Lorge, Teachers College, Columbia University
Staircase Methods.
Dr. Philip J. McCarthy, Cornell University
Inefficient Statistics.
Dr. Frederick Mosteller, Harvard University

Discussion: Dr. Jack W. Dunlap, Psychological Corporation
Dr. Leon Festinger, Massachusetts Institute of Technology
Dr. William E. Kappauf, Princeton University
Dr. Joseph Zubin, New York Psychiatric Institute Hospital

P. S. DWYER,
Secretary.

REPORT ON THE SAN DIEGO MEETING OF THE INSTITUTE

The first Western Regional meeting of the Institute of Mathematical Statistics was held in San Diego, California, June 17–19, 1947, jointly with the American Association for the Advancement of Science. The meeting was attended by 53 persons, including the following 31 members of the Institute:

G. A. Baker, Joseph Berkson, Z. W. Birnbaum, H. C. Carver, Harald Cramér, J. R. Crawford, Dorothy Cruden, W. J. Dixon, Robert Dorfman, M. W. Eudey, Evelyn Fix, John Gurland, J. L. Hodges, Jr., J. M. Howell, H. M. Hughes, E. S. Keeping, E. L. Lehmann, R. H. Lien, F. J. Massey, G. F. McEwen, Frederick Mosteller, S. W. Nash, Jerzy Neyman, Kathryn B. Rolfe, Henry Scheffé, Herbert Solomon, C. M. Stein, Zenon Szatrowski, H. M. Walker, J. D. Williams, Zivia S. Wurtele.

The afternoon session on June 17 was a joint meeting with the Group of Former Operations Analysts. The following program was presented under the chairmanship of Col. Roscoe C. Wilson:

Topic: *Statistical Problems in Operations Analysis.*
Papers: *Engineering and Statistics at the Pacific Front in World War II.*
Roger Wilkinson, Bell Telephone Laboratories, New York City.
Present Organization and Activities of Operations Analysis.
Leroy A. Brothers, Operations Analysis, Asst. Chief of Air Staff-3, Washington, D. C.
Statistical Evidence of Bomb Release Malfunctions.
Mark W. Eudey, University of California, Berkeley.
Study of Effectiveness of Certain Bombs Used Against German Industrial Targets.
J. Neyman, University of California, Berkeley.

The morning session on June 18 was presented with Professor Alva R. Davis as chairman, and the program was as follows:

Topic: *Statistical Problems in Biology.*
Papers: *A Mathematical Model of the Relation between White and Yolk Weights of Birds' Eggs.*
G. A. Baker, University of California, Davis.
Statistical Analysis for a New Procedure in Sensitivity Experiments.
W. J. Dixon, University of Oregon, and A. M. Mood, Iowa State College.
The Relation of Inbreeding to Calf Mortality.
P. W. Gregory, University of California, Davis.
Cooperative Field Trials.
P. A. Minges, University of California, Davis.
Population Genetics.
N. H. Horowitz, California Institute of Technology.
Statistical Problems in Assessing Methods of Medical Diagnosis, with Particular Reference to X-Ray Technique.
J. Yerushalmy, United States Public Health Service, Washington, D. C.
Discussion: J. Neyman, University of California, Berkeley.

Professor John W. Miles was chairman of the afternoon session on June 18, which was a joint session with the California Section of the American Society for Quality Control. The following papers were presented:

Topic: *Industrial Applications of Statistics.*
Papers: *Operating Characteristics of Average and Range Charts.*
         Henry Scheffé, University of California, Los Angeles.
         Sampling Inspection by Variables.
         Herbert Solomon, Stanford University.
         Some Exact Numerical Results for Sequential Acceptance Sampling by Attributes.
         Mark W. Eudey, University of California, Berkeley.
         Choice of Inspection Stringence in Acceptance Sampling by Attributes.
         Joseph L. Hodges, University of California, Berkeley.
         Widening Tolerances for Closer Fitting Parts.
         Edmond E. Bates, Quality Engineering Consultants, Los Angeles.
         Discussion: Russell O'Neill, University of California, Los Angeles.
         Re-establishing Operator Responsibility for Quality Control.
         Wyatt H. Lewis, General Electric Company, Ontario, California.
         Discussion: William B. Rice, Plomb Tool Company, Los Angeles.
         The Application of Learning Curves to Industrial Planning.
         James R. Crawford, Wright Field, Dayton, Ohio.

The Wednesday evening session was under the chairmanship of Professor George Beadle, California Institute of Technology, with the following program:

Topic: *Statistical Problems in Genetical Studies in Chickens.*
Paper: *Rate of Genetic Gain in Egg Production in Progeny-tested Flocks as a Function of the Interval between Generations.*
         Everett R. Dempster and I. Michael Lerner, University of California, Berkeley.

On Thursday morning, June 19, there was a joint session with the Western Psychological Association. Professor Helen Walker of Columbia University was chairman. The program was as follows:

Topic: *Statistical Problems in Psychology.*
Papers: *Statistical Criteria of the Effectiveness of Selective Procedures.*
         R. F. Jarrett, University of California, Berkeley.
         Unsolved Statistical Problems Arising in Psychological Measurements.
         Helen Walker, Columbia University.
         Cost Utility Curves as a Means of Assessing Batteries of Tests.
         Joseph Berkson, Mayo Clinic.
         Approaches to Univocal Factor Scores.
         J. P. Guilford, University of Southern California.

The afternoon session on June 19 was under the chairmanship of Professor Harald Cramér of Stockholm, Sweden, and offered the following program:

Topic: *Theory of Statistics and its Applications to Astronomy.*
Papers: *Random Variables with Comparable Peakedness.*
         Z. W. Birnbaum, University of Washington.
         Distributions which Lead to Regressions Representable by Polynomials.
         Evelyn Fix, University of California, Berkeley.
         Optimum Tests of Composite Hypotheses with One Constraint.
         Erich L. Lehmann, University of California, Berkeley.
         Estimation of a Distribution Function by Confidence Limits.
         Frank J. Massey, Jr., University of California, Berkeley.
         A Note on Sequential Confidence Sets.
         Charles Stein, Columbia University.

Certain Types of Statistical Problems in Astronomy.
Robert J. Trumpler, University of California, Berkeley.
Basic Concepts of the Theory of Statistics in Relation to Certain Problems of Astronomy.
J. Neyman, University of California, Berkeley.
A Note on the Problem of Binary Stars.
Elizabeth L. Scott, University of California, Berkeley.
Explicit Solution of the Problem of Fitting a Straight Line when both Variables are Subject to Error for the Case of Unequal Weights. (By title)
Elizabeth L. Scott, University of California, Berkeley.
Power Function of the Analysis of Variance and Covariance Test of a Normal Bivariate Population. (By title)
Way Ming Chen, University of California, Berkeley.
Unbiased Estimates with Minimum Variance. (By title)
Charles Stein, Columbia University.
Sufficient Statistics and a System of Partial Differential Equations. (By title)
Erich L. Lehmann, University of California, Berkeley, and Henry Scheffé, University of California, Los Angeles.

On Wednesday evening, June 18, at 6 o'clock, there was a dinner for members and guests, at the Hotel San Diego.

P. S. DWYER
Secretary

ON OPTIMUM TESTS OF COMPOSITE HYPOTHESES WITH ONE CONSTRAINT[1]

By E. L. Lehmann

University of California, Berkeley

Summary. This paper is concerned with optimum tests of certain composite hypotheses. In section 2 various aspects of a theorem of Scheffé concerning type B_1 tests are discussed. It is pointed out that the theorem can be extended to cover uniformly most powerful tests against a one-sided set of alternatives. It is also shown that the method for determining explicitly the optimum test region may in certain cases be reduced to a simple formal procedure. These results are used in section 3 to obtain optimum tests for the composite hypothesis specifying the value of the circular serial correlation coefficient in a normal distribution. A surprising feature of this example is the fact that for the simple hypothesis obtained by specifying values for the nuisance parameters no test with the corresponding optimum properties exists.

In section 4 the totality of similar regions is obtained for a large class of probability laws which admit a sufficient statistic. Some composite hypotheses concerning exponential and rectangular distributions are treated in section 5. It is proved that the likelihood ratio tests of these hypotheses have various optimum properties.

1. Introduction. In developing tests for a class of hypotheses three phases may be distinguished. First, tests are obtained which are intuitively appealing; next, it is shown that these tests have certain attractive features; finally, it is proved that they are "best possible" tests.

In dealing with parametric hypotheses, the likelihood ratio principle is frequently used to obtain a reasonable test. For many of the tests so derived for normal and exponential distributions, the question of bias has been investigated. In most cases unbiasedness has been established; in the other cases, usually a test based on the same criterion but with the boundaries shifted, can be proved to be unbiased. Other desirable properties which likelihood ratio tests have been shown to possess, relate to the asymptotic behaviour of these tests as the sample sizes tend to infinity. An interesting problem which does not seem to have been treated is the question of admissibility of likelihood ratio tests, a test being admissible if its power can not be improved upon uniformly by any other test of the same level of significance.

Investigations of optimum tests of composite hypotheses have been carried through for many hypotheses concerning normal distributions. When the hypothesis specifies the value of one parameter (hypothesis with one constraint), uniformly most powerful one-sided and type B_1 (uniformly most powerful un-

---

biased) tests have been obtained. When the number of constraints is larger than one, not so much can be expected. It has been shown for some of the tests in this class that they have maximum average power uniformly over a family of surfaces in the parameter space, or that they are uniformly most powerful with respect to the subclass of tests whose power depends only on some function of the parameters. (All optimum properties mentioned are relative only to the class of all similar regions. This will be so throughout the paper and will usually not be stated explicitly).

Two methods for finding uniformly most powerful or uniformly most powerful one-sided regions and type B_1 tests, if they exist are known. Neyman and Pearson [1] developed a method for determining all similar regions, and applied it to obtain uniformly most powerful one-sided tests of certain hypotheses. Neyman [2, 3] extended the method to obtain, for certain hypotheses, the class of all bisimilar (unbiased similar) regions, and Scheffé [4], developing the method further, proved the existence of type B_1 tests for an important class of hypotheses.

A different method for obtaining all similar and bisimilar regions was devised by P. L. Hsu and was used by him and other writers to prove various optimum properties of the likelihood ratio tests for the general linear hypothesis, of Hotelling's T^2 and of other tests [5, 6, 7, 8].

In the present paper we are concerned with applications of these two methods to composite hypotheses with one constraint. However, the applicability is not so restricted. In fact, the second method has been used mainly in connection with composite hypotheses with many constraints, and the author believes it to be suitable also for deriving optimum classification procedures. An essential restriction of both methods seems to be that a set of sufficient statistics must exist with respect to the parameters involved: with respect to the nuisance parameters so that all similar regions can be found, with respect to the parameters specified by the hypothesis so that there exists a best of all similar regions.

Extensions of the existing theory based on the first method are obtained in section 2, and the theory is applied in section 3 to a hypothesis concerning a multivariate normal distribution. Sections 4 and 5 are concerned with applications of the second method to problems to most of which the earlier method is not applicable, in particular to hypotheses concerning exponential and rectangular distributions, hitherto only treated from the likelihood ratio point of view.

2. On the theory of optimum tests.

2.1 *One-sided tests.* In an interesting paper [4], Scheffé determined the type B and type B_1 tests of a certain class of composite hypotheses specifying the value θ_0 of a parameter θ in the presence of nuisance parameters.

Scheffé's results can, in an obvious way, be extended to cover one-sided sets of alternatives. To show this, consider the method used in [4]. Under certain assumptions all tests[2] are found which satisfy the two conditions:

---

[2] The terms "the test w" and "the region [of rejection] w " will be used interchangeably.

(a) The power function β_w at θ_0 has a preassigned value ϵ (the level of significance), independent of the nuisance parameters;

(b) the power function at θ_0 has derivative 0. (Condition of unbiasedness). Then that test w_0 is determined for which, of all those satisfying (a) and (b),

(c) the second derivative at θ_0, $\beta_w''(\theta_0)$, is as large as possible.

By definition w_0 is a type B test. Under a certain additional assumption (this is the convexity assumption $\dfrac{\partial^2 g}{\partial y_1^2} > 0$ of Scheffé's Theorem 2) it is shown that of all tests satisfying (a) and (b), w_0 has maximum power against all alternatives, i.e. is of type B_1.

If now we want to maximize the power against only the one-sided set of alternatives, $\theta > \theta_0$, we determine that test w_1 of all those satisfying (a), for which

(d) the first derivative at θ_0, $\beta_w'(\theta_0)$, is as large as possible.

Under a certain additional assumption (in Scheffé's notation this would be the monotonicity assumption $\dfrac{\partial g}{\partial y_1} > 0$) it can then be shown that of all tests satisfying (a), w_1 has maximum power against all alternatives $\theta > \theta_0$, (it also has minimum power against all alternatives $\theta < \theta_0$), i.e. w_1 is uniformly most powerful against alternatives $\theta > \theta_0$. We shall not carry through the discussion in detail since Scheffé's argument applies step by step, with only the obvious changes.

2.2 *Determination of the boundaries.* Let $X_1, \cdots, X_n$ be n random variables with a joint probability density function p, depending on parameters θ_1 and $\theta = (\theta_2, \cdots, \theta_l)$. We shall denote the probability density function of a set of random variables $X_1, \cdots, X_n$ whose distribution depends on a parameter θ by $p(x_1, \cdots, x_n \mid \theta)$ or simply by $p(x_1, \cdots, x_n)$ when the dependence on θ is clear from the context. The set of points $(x_1, \cdots, x_n)$ for which

$$p(x_1, \cdots, x_n \mid \theta)$$

is positive we shall denote by $W_+(\theta)$.

Let

$$(2.1) \qquad \varphi_i(x_1, \cdots, x_n) = \frac{\partial}{\partial \theta_i} \log p(x_1, \cdots, x_n \mid \theta_1, \theta) \mid_{\theta_1 = \theta_1^0}, \qquad (i = 1, \cdots, l),$$

and let the random variable Φ_i be defined by

$$(2.2) \qquad \Phi_i = \varphi_i(X_1, \cdots, X_n).$$

Then for testing the hypothesis $H: \theta_1 = \theta_1^0$, under the assumptions stated by Scheffé, the type B_1 test w_0 is defined by the inequalities

$$(2.3) \qquad\qquad \varphi_1 < k_1, \ > k_2 \qquad\qquad (k_1 < k_2)$$

where k_1, k_2 depend on θ_1^0, θ, φ_2, $\cdots$, φ_l and are determined by the two equations[3]

$$(2.4) \qquad \int_{k_1}^{k_2} \varphi_1^s \, p(\varphi_1, \cdots, \varphi_l) \, d\varphi_1 = (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1).$$

---

[3] Although k_1 and k_2 may depend on θ, w_0 is independent of θ, as was shown in [4].

The equations (2.3) and (2.4) are not suitable for the determination of the boundary of w_0. The variables have to be transformed so as to obtain for w_0 an expression from which the calculation of the boundaries becomes feasible, (cf. [9]). This part of the work may be formalized in the following theorem.

THEOREM 1. *Let*

$$(2.5) \quad \begin{aligned} U &= f(\Phi_1, \Phi_2, \cdots, \Phi_l) \\ V_i &= g_i(\Phi_2, \cdots, \Phi_l), \end{aligned} \qquad (i = 2, \cdots, l),$$

be a system of functions, continuously differentiable and with non-vanishing Jacobian almost everywhere, and such that

(i) *U is a linear function of Φ_1*

$$(2.6) \qquad U = a\Phi_1 + b$$

with coefficients which may depend on $\Phi_2, \cdots, \Phi_l$ and such that[4] $a(\Phi_2, \cdots, \Phi_l) > 0$,

(ii) *it is possible to solve for $\Phi_2, \cdots, \Phi_l$ in terms of the V's,*

(iii) *under the hypothesis H, U is distributed independently of*

$$V = (V_2, \cdots, V_l).$$

Then the region w_0 is equivalent to the region

$$(2.7) \qquad u < c_1, \ > c_2 \qquad (c_1 < c_2)$$

where c_1, c_2 are determined by

$$(2.8) \qquad \int_{c_1}^{c_2} u^s p(u)\, du = (1 - \epsilon) \int_{-\infty}^{\infty} u^s p(u)\, du \qquad (s = 0, 1).$$

PROOF.

$$(2.9) \quad \begin{aligned} p(\varphi_1, \varphi_2, \cdots, \varphi_l) &= p(u, v_2, \cdots, v_l) \cdot \left| \frac{\partial(u, v_2, \cdots v_l)}{\partial(\varphi_1, \cdots, \varphi_l)} \right| \\ &= p(u) \cdot p(v_2, \cdots, v_l) \frac{\partial u}{\partial \varphi_1} \cdot \left| \frac{\partial(v_2, \cdots, v_l)}{\partial(\varphi_2, \cdots, \varphi_l)} \right|. \end{aligned}$$

But

$$(2.10) \quad \begin{aligned} u &= a(\varphi_2, \cdots, \varphi_l) \cdot \varphi_1 + b(\varphi_2, \cdots, \varphi_l) \\ &= \alpha(v_2, \cdots, v_l) \cdot \varphi_1 + \beta(v_2, \cdots, v_l) \end{aligned}$$

so that (2.4) reduces to

$$(2.11) \quad \int_{c_1(v_2, \cdots, v_l)}^{c_2(v_2, \cdots, v_l)} \left(\frac{u - \beta}{\alpha} \right)^s p(u) p(v_2, \cdots, v_l)\, du$$
$$= (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1)$$

---

[4] A similar theorem holds when we assume $a(\Phi_2, \ldots, \Phi_l) < 0$.

and hence to

$$(2.12) \qquad \int_{c_1(v_2,\cdots,v_l)}^{c_2(v_2,\cdots,v_l)} u^s\, p(u)\, du \;=\; (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1)$$

which shows c_1 and c_2 to be independent of the v's. Also obviously (2.3) transforms into (2.7) which completes the proof.

If U is such that its distribution (when $\theta_1 = \theta_1^0$) is independent of θ, c_1 and c_2 of theorem 1 will depend only on the data of the problem: ϵ, n, θ_1^0. However, the existence of constants c_1 and c_2 satisfying (2.8) still has to be proved. We may show more generally the existence of k_1 and k_2 satisfying (2.4). A proof is immediately supplied by an argument which was used by Neyman [10] and Wald [11] to prove the existence of type Λ tests, and which may be stated in the following

LEMMA. *Let* $0 < \alpha < 1$, *let* $f(x) \geq 0$ *and* $\int_{-\infty}^{\infty} x^s f(x)\, dx < \infty$ *for* $s = 0, 1$. *Then there exist* A, B *such that*

$$(2.13) \qquad \int_A^B x^s f(x)\, dx \;=\; \alpha \int_{-\infty}^{\infty} x^s f(x)\, dx \qquad (s = 0, 1).$$

3. Testing for circular serial correlation in a normal population. We now apply the results of the previous section to obtain the optimum tests (i.e. uniformly most powerful against the one-sided set of alternatives, type B_1 in the two-sided case) for the hypothesis specifying the value of the circular serial correlation coefficient in the normal population considered by Dixon [12]. (For the literature on testing for non-circular serial correlation in normal populations cf. [12]).

We assume

$$(3.1) \quad p(x_1, \cdots, x_n) = \frac{1 - \delta^n}{(\sqrt{2\pi}\sigma)^n} \exp\left[- \frac{1}{2\sigma^2} \sum_{i=1}^{n} [(x_i - \xi) - \delta(x_{i+1} - \xi)]^2 \right]$$

where $x_{n+1} = x_1$ and $|\delta| < 1$, and we test the hypothesis $\delta = \delta_0$. For testing purposes only the value $\delta_0 = 0$ is of interest presumably, however, the family of tests for arbitrary δ_0 is required for estimating δ by means of confidence intervals, and therefore the more general hypothesis is considered.

Making a transformation in one of the parameters we write

$$p(x_1, \cdots, x_n)$$
$$(3.2)$$
$$= C(\delta, \alpha) \exp\left[\alpha \left[(1 + \delta^2) \sum_{i=1}^{n} (x_i - \xi)^2 - 2\delta \sum_{i=1}^{n} (x_i - \xi)(x_{i+1} - \xi) \right] \right]$$

where in the notation of the previous section $\theta_1 = \delta$, $\theta_2 = \alpha$, $\theta_3 = \xi$.

THEOREM 2. *For testing the hypothesis* $\delta = \delta_0$ *for the distribution* (3.2)
(a) *the type* B_1 *test exists and is given by*

$$(3.3) \qquad r < r_1, \; > r_2$$

where

$$(3.4) \qquad r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \; ; \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

and where r_1 and r_2 are determined by

$$(3.5) \qquad \int_{r_1}^{r_2} \left(\frac{r}{1 + \delta_0^2 - 2\delta_0 r} \right)^s p(r) \, dr = (1 - \epsilon) \int_{-\infty}^{\infty} \text{same} \qquad (s = 0, 1).$$

(b) *the uniformly most powerful similar region for testing H against the alternatives $\delta > \delta_0$ exists and is given by*

$$(3.6) \qquad\qquad\qquad r > r'$$

where r' is determined by

$$(3.7) \qquad\qquad \int_{-\infty}^{r'} p(r) \, dr = (1 - \epsilon) \int_{-\infty}^{\infty} p(r) \, dr.^{5}$$

PROOF. We compute

$$\varphi_1 = C_1(\delta_0, \alpha) + 2\alpha[\delta_0 \Sigma(x_i - \xi)^2 - \Sigma(x_i - \xi)(x_{i+1} - \xi)]$$

$$(3.8) \quad \varphi_2 = C_2(\delta_0, \alpha) + (1 + \delta_0^2)\Sigma(x_i - \xi)^2 - 2\delta_0\Sigma(x_i - \xi)(x_{i+1} - \xi)$$

$$\varphi_3 = -2n\alpha(1 - \delta_0^2)(\bar{x} - \xi).$$

There is no difficulty in checking the conditions of Scheffé's theorems [4].

Next we apply Theorem 1 of the previous section, and define

$$V_2 = (1 + \delta_0^2)\Sigma(X_i - \bar{X})^2 - 2\delta_0\Sigma(X_i - \bar{X})(X_{i+1} - \bar{X})$$

$$(3.9) \qquad V_3 = \bar{X} - \xi$$

$$U = \frac{\Sigma(X_i - \bar{X})(X_{i+1} - \bar{X})}{V_2}.$$

Conditions (i) and (ii) of Theorem 1 are easily seen to be satisfied. To show that U is independent of $V = (V_2, V_3)$ we employ arguments which have recently been used by various authors in a number of similar problems (cf. [13, 14, 15]).

It is seen that an orthonormal transformation exists:

$$X_1, \cdots, X_n \to Y_1, \cdots, Y_n$$

such that

$$\sqrt{n}\bar{X} = Y_1$$

$$(3.10) \qquad \sum_{i=1}^{n} (X_i - \bar{X})(X_{i+1} - \bar{X}) = \sum_{i=2}^{n} \lambda_i Y_i^2$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=2}^{n} Y_i^2.$$

---

[5] A corresponding result holds for the other one-sided case.

Under H the Y's are distributed with probability density

$$(3.11) \quad p(y_1, \cdots, y_n) = C(\delta_0, \alpha) \exp\left[\alpha\left[k(y_1 - \sqrt{n}\,\xi)^2 + \sum_{i=2}^{n} \mu_i y_i^2\right]\right]$$

where $k, \mu_2, \cdots, \mu_n$ depend on δ_0 and where the μ's are all positive. Introducing new variables

$$(3.12) \qquad\qquad Z_i = \sqrt{\mu_i}\, Y_i, \qquad\qquad (i = 2, \cdots, n),$$

and, then, generalized polar coordinates in the space of the Z's,

$$(3.13) \qquad\qquad R = \sqrt{\sum_{i=2} Z_i^2}, \qquad \Psi_1, \cdots, \Psi_{u-2}$$

we see that Y_1, R and $\Psi_1, \cdots, \Psi_{n-2}$ are completely independent. Also

$$V_2 = R^2, \qquad V_3 = \frac{1}{\sqrt{n}}(Y_1 - \xi)$$

while U, being homogeneous of degree 0 in the Z's, is a function of the Ψ's only. This proves that U, V_2 and V_3 are completely independent. The type B_1 test of H is therefore given by

$$(3.14) \quad u = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(x_{i+1} - \bar{x})}{(1 + \delta_0^2)\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2\delta_0 \displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(x_{i+1} - \bar{x})} < c_1, > c_2$$

where c_1 and c_2 are determined by

$$(3.15) \qquad \int_{c_1}^{c_2} u^s p(u)\, du = (1 - \epsilon)\int_{-\infty}^{\infty} u^s pu()\, du \qquad\qquad (s = 0, 1).$$

We still have to show that this test is equivalent to the one defined by (3.3) and (3.5). For $\delta_0 = 0$ this is trivial. Let us assume $\delta_0 < 0$. (The other case goes through similarly.) The inequality $u < c_1$ is equivalent to

$$(3.16) \qquad (1 + 2\delta_0 c_1)\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x}) < (1 + \delta_0^2)\Sigma(x_i - \bar{x})^2$$

and hence to

$$(3.17) \qquad \frac{\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x})}{\Sigma(x_i - \bar{x})^2} < w_1$$

provided $1 + 2c_1\delta_0 > 0$. Suppose $1 + 2c_1\delta_0 \leq 0$, i.e. $c_1 \geq -\dfrac{1}{2\delta_0}$. Then[6]

$$(3.18) \quad P\{U < c_1\} \geq P\left\{U < -\frac{1}{2\delta_0}\right\} = P\{0 < \Sigma(X_i - \bar{X})^2\} = 1$$

---

[6] We denote the probability of an event A by $P\{A\}$.

i.e. $P\{U < c_1\} = 1$ which would contradict (3.15). Similarly if $1 + 2c_2\delta_0 \leq 0$ we would have $P\{U > c_2\} = 0$ and hence our test would be one-sided and therefore not unbiased. The inequalities $u < c_1$, $> c_2$ are thus equivalent to the inequalities $r < r_1$, $> r_2$ and since

$$u = \frac{r}{1 + \delta_0^2 - 2\delta_0 r},$$

(3.5) also follows.

The existence of type B_1 and uniformly most powerful one-sided tests of the hypothesis H is rather surprising. For when α and ξ are assumed known, neither the type A_1 test nor the uniformly most powerful one-sided test of the simple hypothesis H': $\delta = \delta_0$ exists. This is easily seen by determining the most powerful and the most powerful unbiased test against a specific alternative δ_1 for the hypothesis H' in the population

$$(3.19) \quad p(x_1, \cdots, x_n) = \frac{1 - \delta^n}{(\sqrt{2\pi})^n} \exp\left[-\tfrac{1}{2}[(1 + \delta^2)\Sigma x_i^2 - 2\delta\Sigma x_i x_{i+1}]\right].$$

The distribution of the criterion R was obtained by R. L. Anderson [16] (see also [17]) for the case $\delta = 0$. Madow [15] using Anderson's result found the distribution for arbitrary δ. (Approximations to the distribution have been studied by various authors; for the literature on this cf. [18]. Recently Hsu [19] obtained an asymptotic expansion.) A direct derivation for arbitrary δ may be based on the following theorem of Cramér, which was communicated to the author by Dr. P. L. Hsu.

THEOREM 3. (Cramér)[7]. *If X, Y are two random variables, (not necessarily independent), $Y > 0$, then*

$$(3.20) \qquad P\left\{\frac{X}{Y} \leq x\right\} = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\varphi_x(t) - \psi(t)}{it}\, dt$$

where φ_x and ψ are the characteristic functions of $X - xY$ and Y respectively, provided

$$(3.21) \qquad \int_{-\infty}^{\infty} \left|\frac{\varphi_x(t) - \psi(t)}{t}\right| dt < \infty.$$

THEOREM 4. *If*

$$(3.22) \quad \begin{aligned} p(x_1, \cdots, x_n) &= \frac{1 - \delta^n}{(\sqrt{2\pi}\,\sigma)^n} \\ &\cdot \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i - \xi) - \delta(x_{i+1} - \xi)]^2\right], \qquad (x_{n+1} = x_1) \end{aligned}$$

---

[7] Differentiated forms of the theorem were given by R. C. Geary [*Jour. Roy. Stat. Soc.* Vol. 107 (1944) p. 56] and H. Cramér [Exercise 6 on p. 317 of *Mathematical Methods of Statistics.* Princeton Univ. Press (1946)].

and if

$$(3.23) \qquad R = \frac{\sum\limits_{i=1}^{n} (X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum\limits_{i=1}^{n} (X_i - X^2)},$$

then

$$(3.24) \qquad P\{R > r\} = \frac{2^{n+1/2}}{n} \frac{1 - \delta^n}{(1 - \delta)(1 + \delta^2 - 2\delta r)}$$

$$\cdot \sum_j \frac{(-1)^{j+1}(\lambda_j - r)^{n-3/2} \sin\frac{j\pi}{n} \sin\frac{2j\pi}{n}}{1 + \delta^2 - 2\delta\lambda_j}$$

where the summation is extended over all integer j, $1 \le j \le \dfrac{n}{2}$, for which $\lambda_j > r$, and where

$$(3.25) \qquad \lambda_j = 2 \cos\frac{2j\pi}{n}.$$

The proof of this theorem from Theorem 3 is straightforward and only will be indicated here. If X and Y denote the numerator and denominator of R respectively, the characteristic functions of Y and $X - rY$ may be obtained by the method of circulants (cf. [12, 17]). The integral on the right hand side of (3.20) is then easily evaluated by the theory of residues when n is odd. In the case that n is even, the integrand has two branchpoints, one in the lower and one in the upper half plane. These may be separated, and then again the method of residues may be applied.

4. Similar regions. The problem of finding all regions similar to the sample space with respect to a parameter θ was solved by Neyman and Pearson [1] for a certain class of probability laws. In a later paper Neyman proved ([20] proposition IX) that if there exists a sufficient statistic T for a parameter θ, then w is similar with respect to θ if it has the following structure: For the intersection $w(t)$ of w with the surface $T = t$, the relative probability of $w(t)$ given $T = t$ has a constant value independent of t. We shall show in this section that for a large class of probability laws which admit a sufficient statistic for θ the regions with the above structure are the only ones that are similar with respect to θ.

We consider samples from a univariate distribution and we distinguish three cases as one, both or neither of the extremes of the range of the distribution depend on the parameter θ. For the first of these cases (cf. Pitman [21]) we consider samples from a distribution with probability density

$$(4.1) \qquad\qquad p(x) = \frac{f(x)}{g(\theta)}, \qquad k(\theta) \le x \le c,$$

where $k(\theta)$ is a strictly monotone continuous function of θ and where c may be infinite. Introducing a new parameter $\delta = k(\theta)$ the distribution of a sample from (4.1) is given by

$$(4.2) \qquad\qquad p(x_1, \cdots, x_n) = \frac{f(x_1) \cdots f(x_n)}{b(\delta)}, \qquad \delta \le x_i \le c.$$

To obtain the totality of regions w similar with respect to δ let us denote by $W_1, \cdots, W_n$ the portions of the sample space where the smallest of the x's is $x_1, \cdots, x_n$ respectively. For any region w denote by w_k' the intersection of w with W_k. Consider a transformation carrying $W_2, \cdots, W_n$ into W_1, letting $y_1 = \min(x_1, \cdots, x_n)$ and letting in W_k :

$$(4.3) \quad y_2 = x_1, \ y_3 = x_2, \cdots, \ y_k = x_{k-1}, \ y_{k+1} = x_{k+1}, \cdots, \ y_n = x_n.$$

Denote by w_k the image of w_k' under this transformation. The condition that w be similar with respect to δ,

$$(4.4) \qquad\qquad \int_w \frac{f(x_1) \cdots f(x_n)}{b(\delta)} \, dx_1 \cdots dx_n \overset{(\delta)}{\equiv} \epsilon,$$

may be written in the form

$$(4.5) \quad \int_\delta^c \frac{f(y_1)}{b(\delta)} \left\{ \sum_{k=1}^n \int_{w_k(y_1)} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n \right\} dy_1$$
$$\overset{(\delta)}{\equiv} n\epsilon \int_\delta^c \frac{f(y_1)}{b(\delta)} \left\{ \int_{(W(y_1))} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n \right\} dy_1$$

where $W(y_1)$ denotes the region $y_1 \le y_i \le c, \ (i = 2, \cdots, n)$, that is, the region of variation of $y_2, \cdots, y_n$ given y_1, and where $w_k(y_1)$ denotes the region of variation of $y_2, \cdots, y_n$ given y_1 and w_k. From (4.5) we obtain

$$(4.6) \qquad\qquad \frac{1}{b(\delta)} \int_\delta^c f(y_1)\psi(y_1) \, dy_1 \overset{(\delta)}{\equiv} 0$$

where

$$(4.7) \quad \psi(y_1) = \sum_{k=1}^n \int_{w_k(y_1)} f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n$$
$$- n\epsilon \int_{y_1}^c \cdots \int_{y_1}^c f(y_2) \cdots f(y_n) \, dy_2 \cdots dy_n.$$

But (4.6) implies

$$(4.8) \qquad\qquad \psi(y_1) = 0 \text{ almost everywhere}$$

and since we can only determine w up to a set of measure 0, we may omit the qualification in (4.8). Therefore a necessary and sufficient condition for w to be similar is

(4.9) $\qquad \dfrac{1}{n\left[\displaystyle\int_{y_1}^{a} f(y)\,dy\right]^{n-1}} \sum_{k=1}^{n} \int_{w_k(y_1)} f(y_2) \cdots f(y_n)\,dy_2 \cdots dy_n = \epsilon$

for all y_1.

To see more clearly the structure of these regions, let us take $n = 2$. Equation (4.9) states that on each of the broken lines of Fig. 1 the relative probability of $w = w_1' + w_2'$ given $Y_1 = y_1$ is ϵ, where the decomposition of this probability into its two components may vary with y_1.
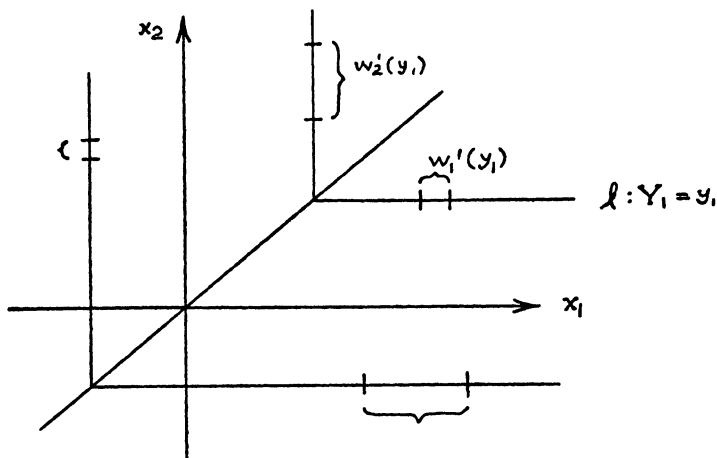

FIG. 1

In general equation (4.9) states that on each hyperplane $Y_1 = y_1$ the relative probability of w is independent of y_1. Since $Y_1 = \min (X_1, \cdots, X_n)$ is a sufficient statistic for θ, Neyman's theorem in this case does give all similar regions.

Next let us consider the case where both extremes of the range of the distribution depend on the parameter. We shall assume (cf. [21]) that $X_1, \cdots, X_n$ are distributed with probability density

(4.10) $\qquad\qquad p(x) = \dfrac{f(x)}{g(\theta)} \quad \text{in} \quad \theta \leq x \leq b(\theta)$

where b is a strictly decreasing continuous function over an interval $[-\infty, b(-\infty)]$ and where $b[b(-\infty)] = -\infty$. These assumptions insure that there exists a unique number a, $-\infty < a < b(-\infty)$, such that $b(a) = a$.

Denote by W_{ij}, $(i, j = 1, \cdots, n; i \neq j)$, the portion of the sample space where the smallest and the largest of the x's are x_i and x_j respectively. Denote by W_{ij1} and W_{ij2} those portions of W_{ij} where x_i is greater than and less than $b^{-1}(x_j)$ respectively. For any region w denote by w_{ijk}' the intersection of w with

W_{ijk}. Consider a transformation carrying the sample-space into W_{1n}, letting $y_1 = \min(x_1, \cdots, x_n)$, $y_n = \max(x_1, \cdots, x_n)$ and in W_{ij} letting $y_2, \cdots, y_{n-1}$ denote the remaining x's in the order of their subscripts. Next make a transformation carrying W_{1n} into W_{1n1}, letting $z_1 = \max[y_1, b^{-1}(y_n)]$, $z_n = \min[y_1, b^{-1}(y_n)]$ and $z_k = y_k$ for $k = 2, \cdots, n-1$. Denote by w_{ijk} the image of w'_{ijk} in W_{1n1}.

Then Z_n is a sufficient statistic for θ (cf. [21]) and there exist functions f_1, g_1 such that the density of Z_n is given by

$$(4.11) \qquad p(z_n) = \frac{f_1(z_n)}{g_1(\theta)} \quad \text{in} \quad \theta \leq z_n \leq a$$

while the distribution of the remaining Z's given Z_n is independent of θ.

The condition that w be a similar region may now be written, analogously to (4.5), in the form

$$(4.12) \quad \int_a^\theta \frac{f_1(z_n)}{g_1(\theta)} \sum_{i,j,k} \int_{w_{ijk}(z_n)} p(z_1, \cdots, z_{n-1} \mid z_n) \, dz_1 \cdots dz_{n-1} \, dz_n \equiv \epsilon \int_\theta^a \frac{f_1(z_n)}{g_1(\theta)} \, dz_n$$

and hence by the argument which led to (4.6), as

$$(4.13) \qquad \sum_{i,j,k} \int_{w_{ijk}(z_n)} p(z_1, \cdots, z_{n-1} \mid z_n) \, dz_1 \cdots dz_{n-1} = \epsilon \qquad \text{for all } z_n.$$

Thus in this case also Neyman's theorem gives the most general similar region.

For the case that neither extreme of the range of the distribution depends on the parameter θ, it has been shown by various authors [22, 21, 23] under slightly varying assumptions concerning the regularity of the distribution function, that the existence of a sufficient statistic implies

$$(4.14) \qquad p(x \mid \theta) = \exp[P(\theta) + T(x)Q(\theta) + R(x)].$$

This (cf. [10]) is a special case of that for which Neyman and Pearson determined the totality of similar regions, however under the restriction that the moments

of $\Phi = \dfrac{\partial}{\partial \theta} \sum_{i=1}^n \log p(X_i)$ uniquely determine the distribution of Φ. We shall

briefly indicate how this assumption may be avoided.

Let $X_1, \cdots, X_n$ be a sample from (4.14), or, more generally, (this is the case considered by Neyman and Pearson), let $X_1, \cdots, X_n$ be distributed with probability density

$$(4.15) \quad \begin{aligned} &p(x_1, \cdots, x_n) \\ &\qquad = \exp[P(\theta) + u(x_1, \cdots, x_n)Q(\theta) + v(x_1, \cdots, x_n)] \end{aligned}$$

in a sample space W_+ which is independent of θ. We shall assume that the set of values which Q takes on contains at least some interval. Introducing $\delta =$

$-Q(\theta)$ as a new parameter, we shall obtain all regions similar to δ (where the set of values of δ contains an interval) for the distribution[8]

$$
\begin{aligned}
p(x_1 &, \cdots , x_n) \\
(4.16) \\
&= \exp\ [p_1(\delta)\ -\ \delta \cdot u(x_1 ,\ \cdots ,\ x_n)\ +\ v(x_1 ,\ \cdots ,\ x_n)]
\end{aligned}
$$

under the assumption that $\sum_{i=1}^{n} \left(\dfrac{\partial u}{\partial x_i}\right)^2 \neq 0$ except possibly on a set of measure 0.

Let us for a moment assume that there exist functions $f_i(x_i ,\ \cdots ,\ x_n)$, $(i = 2,\ \cdots ,\ n)$, with continuous partial derivatives almost everywhere and such that the transformation

$$
(4.17) \qquad y_1 = u(x_1 ,\ \cdots ,\ x_n); \quad y_i = f_i(x_1 ,\ \cdots ,\ x_n), \qquad (i = 2,\ \cdots ,\ n),
$$

is one to one on W_+ except possibly on a subset of measure 0. Applying this transformation we may write the condition of similarity in the form

$$
\begin{aligned}
(4.18) \\
\int_{-\infty}^{\infty} e^{P_1(\delta)-\delta y_1} &\int_{w(y_1)} f(y_1 ,\ \cdots ,\ y_n)\, dy_2 \cdots dy_n \cdot dy_1 \\
&= \epsilon \int_{-\infty}^{\infty} e^{P_1(\delta)-\delta y_1} \int_{W(y_1)} f(y_1 ,\ \cdots ,\ y_n)\, dy_2 \cdots dy_n \cdot dy_1
\end{aligned}
$$

where $W(y_1)$ denotes the region of variation of $y_2 ,\ \cdots ,\ y_n$ given y_1, and where $w(y_1)$ denotes the region of variation of $y_2 ,\ \cdots ,\ y_n$ given y_1 and w. Furthermore $f(y_1 ,\ \cdots ,\ y_n)$ is independent of δ. From the theory of bilateral Laplace transforms it is known that (4.18) implies that

$$
(4.19) \qquad \int_{w(y_1)} f(y_1 ,\ \cdots ,\ y_n)\, dy_2 \cdots dy_n = \epsilon \int_{W(y_1)} f(y_1 ,\ \cdots ,\ y_n)\, dy_2 \cdots dy_n
$$

which is the desired result.

More generally it may be shown that our assumption concerning $u(x_1 ,\ \cdots ,\ x_n)$ insures the existence of functions f_i, $(i = 2,\ \cdots ,\ n)$, such that under the transformation (4.17) no point $(y_1 ,\cdots ,\ y_n)$ has more than a denumerable infinity of counter images in x-space. Our proof can be modified to cover this case. The argument is similar to that used to obtain equations (4.9) and (4.13) which were also arrived at through many to one transformations.

5. Testing exponential and rectangular distributions. In their fundamental 1928 paper [24] on likelihood ratio tests, Neyman and Pearson discussed various hypotheses relating to normal, exponential and rectangular distributions. Later they and other authors developed a theory of similar and bisimilar regions which made it possible to obtain optimum tests of many composite hypotheses with

---

[8] An assumption that we can solve for θ as a function of δ is not needed since we can determine P_1 (δ) by integrating the density (4.16) over W_+.

one constraint concerning normal populations. This theory however is not applicable to most hypotheses concerning exponential or rectangular distributions. We shall in this section obtain optimum tests of some hypotheses relating to these latter distributions, using the method of the previous section.

Let us first consider a sample $X_1, \cdots, X_n$ from an exponential population, the probability density of the sample being.

$$(5.1) \quad p(x_1, \cdots, x_n) = \frac{1}{a^n} \exp\left[-\frac{1}{a} \sum_{i=1}^{n} (x_i - b) \right] \quad \text{if } x_i > b, \quad (i = 1, \cdots, n)$$

and let us consider the two hypotheses $H_1: a = _0$, $H_2: b = b_0$ where, without loss of generality, we shall take $a_0 = 1$, $b_0 = 0$. The likelihood ratio tests of both these hypotheses were shown to be completely unbiased by Paulson [25]. We shall prove

THEOREM 5. *The likelihood ratio tests of H_1 and H_2 are type B_1 and uniformly most powerful, respectively. The one-sided tests based on the likelihood ratio criterion for H_1 are the uniformly most powerful one-sided similar regions for testing this hypothesis.*

PROOF. In order to simplify the argument we shall give a detailed proof only for the restricted class of tests which are symmetric in the variables $X_1, \cdots, X_n$.

For testing H_1 let us make the following transformation introduced by Sukhatme [26]:

$$(5.2) \quad \begin{aligned} Z_1 &= nY_1 \\ Z_i &= (n - i + 1)(Y_i - Y_{i-1}), \quad (i = 2, \cdots, n), \end{aligned}$$

where Y_i is the ith of the X's in order of magnitude. Then

$$(5.3) \quad p(z_1, \cdots, z_n) = \frac{1}{a^n} \exp\left[-\frac{1}{a}(z_1 - nb) - \frac{1}{a} \sum_{i=2}^{n} z_i \right]$$
$$\text{if} \quad z_1 \geq nb; z_i \geq 0 \quad (i = 2, \cdots, n).$$

We want to determine all regions w which under H are similar to the sample space with respect to b, i.e. all regions w satisfying

$$(5.4) \quad \begin{aligned} \int_w e^{-(z_1 - nb)} \exp\left[-\sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n \, dz_1 \\ = \int_{nb}^{\infty} e^{-(z_1 - nb)} \left\{ \int_{w(z_1)} \exp\left[-\sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n \right\} dz_1 \\ \overset{(b)}{\equiv} \epsilon \overset{(b)}{\equiv} \epsilon \int_{nb}^{\infty} e^{-(z_1 - nb)} \, dz_1 \end{aligned}$$

where $w(z_1)$ denotes the intersection of w with the hyperplane $Z_1 = z_1$. Now (5.4) is equivalent to

(5.5) $$e^{nb} \int_{nb}^{\infty} e^{-z_1} f(z_1) \, dz_1 \overset{(b)}{=} 0$$

where

(5.6) $$f(z_1) = \int_{w(z_1)} \exp\left[- \sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n - \epsilon$$

and this in turn is equivalent to

(5.7) $$f(z_1) = 0 \text{ for all } z_1 .$$

Of all the regions w satisfying (5.7) we want to determine the one which against a specific alternative, say a_1, has maximum power, i.e. for which

(5.8) $$\int_{nb}^{\infty} e^{-(1/a_1)(z_1 - nb)} \int_{w(z_1)} \exp\left[- \frac{1}{a_1} \sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n \, dz_1$$

is as large as possible. We thus see that w will have the desired properties if $w(z_1)$ is determined according to the two conditions

(5.9) $$\int_{w(z_1)} \exp\left[- \sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n = \epsilon$$

and

(5.10) $$\int_{w(z_1)} \exp\left[- \frac{1}{a_1} \sum_{i=2}^{n} z_i \right] dz_2 \cdots dz_n = \text{max}.$$

Hence by the Neyman-Pearson fundamental lemma $w(z_1)$ is the set of points satisfying

(5.11) $$\exp\left[\left(- \frac{1}{a_1} \sum_{i=2}^{n} z_i + \sum_{i=2}^{n} z_i \right) \right] \geq C(a_1, z_1)$$

and therefore according as a_1 is greater or less than 1, $w(z_1)$ is determined by

(5.12) $$\sum_{i=2}^{n} z_i = \sum_{i=1}^{n} [x_i - \min(x_1, \cdots, x_n)] \geq k(a_1, z_1), \text{ or}$$
$$\sum_{i=2}^{n} z_i = \sum_{i=1}^{n} [x_i - \min(x_1, \cdots, x_n)] \leq k'(a_1, z_1).$$

But $\sum_{i=2}^{n} Z_i$ is independently distributed of Z_1 and under H the distribution of $\sum_{i=2}^{n} Z_i$ does not depend on a_1, in fact it is a chi-square distribution with $2n - 2$ degrees of freedom. Thus k and k', as determined by (5.9) are independent of a_1 and the two tests (5.12) are uniformly most powerful one-sided.

Next we consider the more restricted class of unbiased similar regions. For w to be unbiased we must have

$$\frac{d}{da}\left\{\frac{1}{a^n}\int_w \exp\left[-\frac{z_1-nb}{a}\right]\exp\left[-\frac{1}{a}\sum_{i=2}^n z_i\right]dz_1\cdots dz_n\right\}\bigg|_{a=1}$$

$$(5.13)\quad =\int_{nb}^\infty (z_1-nb-n)\exp\left[-(z_1-nb)\right]\int_{w(z_1)}\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1$$

$$+\int_{nb}^\infty \exp\left[-(z_1-nb)\right]\int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1 = 0.$$

The first of the integrals in the middle member equals

$$(5.14)\quad \int_0^\infty (z-n)\,e^{-z}\int_{w(z+nb)}\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz$$

$$=\epsilon\int_0^\infty (z-n)\,e^{-z}\,dz = -(n-1)\epsilon.$$

Therefore

$$(5.15)\quad \int_{nb}^\infty e^{-(z_1-nb)}\int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n\,dz_1$$

$$=(n-1)\epsilon = (n-1)\epsilon\int_{nb}^\infty e^{-(z_1-nb)}\,dz_1$$

or

$$(5.16)\quad \int_{nb}^\infty e^{-z_1}\,g(z_1)\,dz_1 \overset{(b)}{\equiv} 0$$

where

$$(5.17)\quad g(z_1) = \int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n - (n-1)\epsilon.$$

Thus finally the condition of unbiasedness reduces to

$$(5.18)\quad \int_{w(z_1)}\left(\sum_{i=2}^n z_i\right)\exp\left[-\sum_{i=2}^n z_i\right]dz_2\cdots dz_n = (n-1)\epsilon$$

and we seek the region $w(z_1)$ which satisfies (5.9), (5.10) and (5.18).

By the fundamental lemma $w(z_1)$ is given by

$$(5.19)\quad \exp\left[-\frac{1}{a_1}\sum_{i=2}^n z_i\right]\geq \left[C_1(a_1,z_1)\sum_{i=2}^n z_i + C_2(a_1,z_1)\right]\cdot\exp\left[-\sum_{i=2}^n z_i\right]$$

which is equivalent to ·

$$(5.20)\quad \sum_{i=2}^n z_i \leq k_1(a_1,z_1),\ \geq k_2(a_1,z_1)$$

where k_1 and k_2 are determined by (5.9) and (5.18), and are therefore independent of z_1 and a. Thus the region (5.20) which of all unbiased similar regions

maximizes the power against the alternative $a = a_1$ is independent of a_1 and hence is a region of type B_1. This completes the proof since it is easily verified that (5.10) is equivalent to the likelihood ratio test.

The proof for regions which are not necessarily symmetric in the variables follows similarly if instead of the transformation (5.2) one uses a transformation $U_i = f_i(X_1, \cdots, X_n)$ which is one to one and such that $U_1 = Z_1$, $U_2 = \sum_{i=2}^{n} Z_i$. The distribution of $U_3, \cdots, U_n$ is then independent of a and b since U_1, U_2 are a pair of sufficient statistics for these parameters, and the proof carries over step by step.

Next we consider the hypothesis $H_2 : b = 0$, and again we restrict ourselves to regions which are symmetric in the variables, although as before the proof can be modified to cover also nonsymmetric regions.

We first make the transformation to $Z_1, \cdots, Z_n$ given by (5.2). In the $n - 1$ dimensional space of $Z_2, \cdots, Z_n$, we then transform to new variables $U, \Psi_1, \cdots, \Psi_{n-2}$ where $U = \sum_{i=2}^{n} Z_i$ and where the Ψ's are the generalized polar angles. Obviously the distribution of the Ψ's does not depend on a, since they are homogeneous of degree 0 in the Z's. Furthermore the Ψ's are independently distributed of U since the probability density of the Z's is constant over the hyperplanes $U = u$. Thus

(5.21)
$$p(z_1, u, \psi_1, \cdots, \psi_{n-2}) = \frac{K}{a^n} \exp\left[-\frac{z_1 - nb}{a} \right] u^{n-2} e^{-u/a} p(\psi_1, \cdots, \psi_{n-2}).$$

We next introduce new variables

(5.22)
$$V = Z_1 + U \text{ and } T = \frac{Z_1}{Z_1 + U}$$

and find

$$p(v, t, \psi_1, \cdots, \psi_{n-2}) = \frac{K}{a^n} \exp\left[-\frac{v - nb}{a} \right] v^{n-1} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})$$
(5.23)

$$\text{for } v \geq nb, \frac{nb}{v} \leq t \leq 1.$$

For w under H_2 to be similar with respect to a, we must have

(5.24)
$$\int_0^\infty \frac{K}{a^n} \exp\left[-\frac{v}{a} \right] v^{n-1} \int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2}) dt d\psi_1 \cdots d\psi_{n-2} \cdot dv$$
$$= \epsilon \int_0^\infty \frac{K}{a^n} \exp\left[-\frac{v}{a} \right] v^{n-1} dv$$

where $w(v)$ designates the intersection of w with the hyperplane $V = v$, and where $w_0(v)$ denotes the part of $w(v)$ lying between the hyperplanes $t = 0$ and $t = 1$.

Hence the condition of similarity may be written as

$$(5.25) \qquad \int_0^{\infty} \exp\left[-\frac{v}{a}\right] v^{n-1} f(v)\, dv = 0 \qquad\qquad \text{for all } a > 0$$

where

$$(5.26) \qquad f(v) = \int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} - \epsilon.$$

By the uniqueness theorem for Laplace transforms, (5.25) implies $f(v) = 0$ for all $v > 0$, so that the condition of similarity finally reduces to

$$(5.27) \qquad \int_{w_0(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} = \epsilon.$$

Of all similar regions, let us find the one which has maximum power. Obviously we want to include in $w(v)$ all points for which $t < 0$. In addition we want to choose $w_0(v)$ such that

$$(5.28) \qquad \int_{w_b(v)} (1 - t)^{n-2} p(\psi_1, \cdots, \psi_{n-2})\, dt\, d\psi_1 \cdots d\psi_{n-2} = \max$$

where $w_b(v)$ is that part of $w(v)$ in which $\max\left(0, \dfrac{nb}{v}\right) \leq t$.

If, for some alternative b, $w_0(v)$ is contained in $\dfrac{nb}{v} < t < 1$, then $w_b(v)$ and $w_0(v)$ coincide and hence (5.28) attains its maximum value ϵ whatever the position of $w_0(v)$ in $\dfrac{nb}{v} \leq t \leq 1$. If on the other hand $\dfrac{nb}{v}$ is so close to 1 that $\dfrac{nb}{v} \leq t \leq 1$ is too small to contain $w_0(v)$, then (5.28) attains its maximum for any $w_0(v)$ containing $\dfrac{nb}{v} \leq t \leq 1$. There exists therefore a unique $w_0(v)$ which maximizes (5.28) for all values of b and v, namely the region defined by

$$(5.29) \qquad\qquad\qquad C(v) \leq t \leq 1$$

where C is determined by (5.27).

Since under H_2, the statistics V and T are independent, C does not depend on v. The test

$$(5.30) \qquad\qquad\qquad t \leq 0, \qquad \geq C$$

which we have just shown to be uniformly most powerful, is also the likelihood ratio test which completes the proof of the theorem.

We shall finally consider an example of an optimum test in connection with a

rectangular distribution. Let $X_1, \cdots, X_n$ be independently and uniformly distributed over $(a, a + \theta)$, where θ is positive. *For testing the hypothesis* $H: a = a_0$, *the test*

$$(5.31) \qquad \frac{Y_1 - a_0}{Y_n - Y_1} \leq 0, \qquad \geq C$$

where Y_1 and Y_n are the smallest and the largest of the X's respectively, is the uniformly most powerful of all similar regions.

The proof of this goes through very much like that for H_2 in Theorem 5. Without loss of generality we take $a_0 = 0$. Also again, to simplify the proof, we restrict ourselves to regions which are symmetric in the variables. We need the following lemma.

LEMMA. *Let $X_1, \cdots, X_n$ be independently and uniformly distributed over* $(a, a + \theta)$. *Let Y_i denote the ith X in order of magnitude, and let*

$$(5.32) \qquad T_n = Y_n, \; T_k = \frac{Y_k}{Y_{k+1}}, \qquad (= 1, \cdots, n - 1).$$

Then for $a > 0$

$$(5.33) \qquad p(t_1, \cdots, t_n) = \frac{n!}{\theta^n} t_n^{n-1} t_{n-1}^{n-2} \cdots t_2$$

when

$$a \leq t_n \leq a + \theta, \; \frac{a}{t_n \cdot t_{n-1} \cdots t_{k+1}} \leq t_k \leq 1, \qquad (k = 1, \cdots, n - 1).$$

This is easily seen by applying the usual method of Jacobians. The inequalities describing the sample space of the T's are equivalent to the following more convenient ones:

$$(5.34) \quad a \leq t_n \leq a + \theta, \frac{a}{t_n} \leq t_1 t_2 \cdots t_{n-1} \leq 1; t_k \leq 1, \qquad (k = 1, \cdots, n - 1).$$

Let us denote by $w(t_n)$ the intersection of a region w with the hyperplane $T_n = t_n$, and by $w_0(t_n)$ that part of $w(t_n)$ contained in the cylinder $0 \leq t_k \leq 1$, $(k = 1, \cdots, n - 1)$; then we find as a necessary and sufficient condition for w to be similar with respect to θ (assuming H)

$$(5.35) \qquad (n - 1)! \int_{w_0(t_n)} t_{n-1}^{n-2} t_{n-2}^{n-3} \cdots t_2 \, dt_{n-1} \cdots dt_1 = \epsilon.$$

Of all regions satisfying (5.35) we want to find the most powerful one. Let us first consider alternatives $a > 0$. If $w_a(t_n)$ denotes the common part of $w_0(t_n)$ and the region

$$(5.36) \qquad \frac{a}{t_n} \leq t_{n-1} t_{n-2} \cdots t_1 \leq 1,$$

we must choose $w_a(t_n)$ such that

$$(5.37) \qquad \int_{w_a(t_n)} t_{n-1}^{n-2} \cdot t_{n-2}^{n-3} \cdots t_2 \, dt_{n-1} \cdots dt_1 = \max.$$

From this it follows easily that against alternatives $a > 0$ the uniformly best choice for $w_0(t_n)$ is

$$(5.38) \qquad t_1 t_2 \cdots t_{n-1} = \frac{y_1}{y_n} \geq C'(t_n),$$

and since under H, $\frac{Y_1}{Y_n}$ is independently distributed of T_n, $C'(t_n)$ does not depend on t_n.

Consider next alternatives $a < 0$. We include in the region of rejection all points for which $Y_1 \leq 0$. To determine $w_0(t_n)$ we notice that, given $Y_1 > 0$, the X's are uniformly distributed between 0 and $a + \theta$. (Provided $a + \theta > 0$; the case $a + \theta \leq 0$ is trivial). Hence the probability distribution of the T's given $Y_1 > 0$ is

$$(5.39) \qquad p(t_1, \cdots, t_n \mid Y_1 > 0) = \frac{n!}{(a + \theta)^n} t_n^{n-1} \cdots t_2$$

when

$$0 \leq t_n \leq a + \theta, \qquad 0 \leq t_k \leq 1 \qquad \qquad \text{for } k = 1, \cdots, n - 1.$$

Thus

$$(5.40) \qquad \frac{p(t_1, \cdots, t_{n-1} \mid t_n, a < 0, Y_1 > 0)}{p(t_1, \cdots, t_{n-1} \mid t_n, a = 0)}$$

is independent of $t_1, \cdots, t_{n-1}$ and hence the power of w against alternatives $a < 0$ is independent of the choice of $w_0(t_n)$. Therefore the region

$$(5.41) \qquad y_1 \leq 0, \frac{y_1}{y_n} \geq C'$$

is uniformly most powerful against all alternatives. But (5.41) is equivalent to

$$(5.42) \qquad \frac{y_1}{y_n - y_1} \leq 0, \geq C.$$

It is interesting to compare this result with that for the corresponding simple hypothesis. Let H' be the hypothesis: $a = 0$ when the X's are assumed independently and uniformly distributed over $(a, a + 1)$. There exists no uniformly most powerful test of H'; instead the two uniformly most powerful one-sided tests exist. By analogy with the normal case one might then expect for H' that of all tests with symmetric power-functions, there be a uniformly most powerful one. This however is not so: there exist infinitely many admissible tests with symmetric powerfunction.

In this and the previous section we restricted ourselves to problems involving only one nuisance parameter. However, the method applies also to problems involving several nuisance parameters.

In the usual way (cf. [20, 9]) the results of this section may be translated to give optimum sets of confidence intervals for estimating the parameters in question. In this connection it is an open question whether the confidence regions based on the type B_1 tests discussed in section 2 will always be intervals; one would expect this to be the case.

The author wishes to acknowledge his indebtedness to Professor P. L. Hsu for many helpful suggestions.

Added in proof: In a joint paper by Professor Henry Scheffé and the present author which has been submitted to the Proceedings of the National Academy of Sciences, a result is given concerning the existence of certain 1:1 transformations. This result bears on Section 4 of the present paper where a question arises concerning the existence of a 1:1 transformation. The existence of such a transformation is now assured and, as a consequence, the last paragraph of Section 4 has become superfluous.

REFERENCES

[1] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses", *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 231 (1933), p. 289.
[2] J. NEYMAN, "Sur la vérification des hypothèses statistiques composées", *Bull. Soc. Math. France*, Vol. 63 (1935), p. 1.
[3] J. NEYMAN, "On a statistical problem arising in routine analysis and in sampling inspection of mass production", *Annals of Math. Stat.*, Vol. 12 (1941), p. 46.
[4] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint", *Annals of Math. Stat.*, Vol. 13 (1942), p. 280.
[5] P. L. HSU, "Analysis of variance from the power function standpoint", *Biometrika*, Vol. 32 (1941), p. 62.
[6] J. B. SIMAIKA, "On an optimum property of two important statistical tests", *Biometrika*, Vol. 32 (1941), p. 70.
[7] A. WALD, "On the power function of the analysis of variance test", *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.
[8] P. L. HSU, "On the power function of the E^2-test and the T^2-test", *Annals of Math. Stat.*, Vol. 16 (1945), p. 278.
[9] H. SCHEFFÉ, "On the ratio of the variances of two normal populations", *Annals of Math. Stat.*, Vol. 13 (1942), p. 371.
[10] J. NEYMAN, "L'estimation statistique traitée comme un problème classique de probabilité", *Actualités Sci. et Ind. No. 739*, Hermann et Cie, Paris, 1938.
[11] A. WALD, "Notes on the theory of statistical estimation and testing hypotheses", Columbia University.
[12] W. J. DIXON, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), p. 119.
[13] J. VON NEUMANN, "Distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 12 (1941), p. 367.
[14] M. KAC, "A remark on independence of linear and quadratic forms involving independent Gaussian variables", *Annals of Math. Stat.*, Vol. 16 (1945), p. 400.
[15] W. G. MADOW, "Note on the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), p. 308.

[16] R. L. ANDERSON, "Distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 13 (1942), p. 1.

[17] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables", *Annals of Math. Stat.*, Vol. 13 (1942), p. 14.

[18] R. B. LEIPNIK, "Distribution of the serial correlation coefficient in a circularly correlated universe", *Annals of Math. Stat.*, Vol. 18 (1947), p. 80.

[19] P. L. HSU, "On the asymptotic distributions of certain statistics used in testing the independence between successive observations from a normal population", *Annals of Math. Stat.*, Vol. 17 (1946), p. 350.

[20] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability", *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 236 (1937), p. 333.

[21] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy", *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), p. 567.

[22] B. O. KOOPMAN, "On distributions admitting a sufficient statistic", *Trans. Amer. Math. Soc.*, Vol. 39 (1936), p. 399.

[23] D. DUGUÉ, "Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation", *J. École Poly.*, Vol. 3 (1937), p. 305.

[24] J. NEYMAN AND E. S. PEARSON, "On the use and interpretation of certain test criteria for purposes of statistical inference", *Biometrika*, Vol. 20A (1928), p. 175.

[25] E. PAULSON, "On certain likelihood ratio tests associated with the exponential distribution", *Annals of Math Stat.*, Vol. 12 (1941), p. 301.

[26] P. V. SUKHATME, "Tests of significance for samples of the χ^2 population with two degrees of freedom", *Annals of Eugenics*, Vol. 8 (1937), p. 52.

A CORNER TEST FOR ASSOCIATION

By Paul S. Olmstead and John W. Tukey

Bell Telephone Laboratories and Princeton University

1. Summary. This paper proposes a new test (the "quadrant sum") for the association of two continuous variables. Its notable properties are:
(1) Special weight is given to extreme values of the variables.
(2) Computation is very easy.
(3) The test is non-parametric.
Significance levels (for the quadrant sum) are given to the accuracy needed for practical use. To this accuracy they are independent of sample size (see Fig. 1). The generating function of the quadrant sum is given for the null hypothesis (no association = independence). A limiting distribution is deduced and compared with the cases $2n$ = 4, 6, 8, 10, and 14. Extension to higher dimensions and application to serial correlation are discussed.

2. Description of test (even number in sample). We shall describe the test as though a scatter diagram had already been drawn. The possibilities of direct computation from tabular data are indicated by the examples in sections 8 and 9.

In the scatter diagram, draw the two lines, $x = x_m$, $y = y_m$, where x_m is the median of the x-values without regard to the values of y, and y_m is the median of the y-values without regard to the values of x. Think of the four quadrants or corners thus formed as being labelled $+$, $-$, $+$, $-$, in order, so that the upper right and lower left quadrants are positive. Beginning at the right hand side of the diagram, count in (in order of abscissae) along the observations until forced to cross the horizontal median. Write down the number of observations met before this crossing, attaching the sign $+$ if they lay in the $+$ quadrant, and the sign $-$ if they lay in the $-$ quadrant. Repeat this process moving up from below, moving to the right from the left, and moving down from above. The quadrant sum is the algebraic sum of the four terms thus written down. This process is illustrated in Fig. 2, where the black dots represent contributions to the sum, and the dotted lines, crossings.

When there are an even number of pairs (x, y) and no ties, the medians will pass between the points. In this, the simplest case, the distribution of the quadrant sum is known for the hypothesis of no association (that is, of independence), and significance levels are given in Table 1 for the magnitude (absolute value) of the sum. It will be noticed that the sample size does not enter in any important way.

The cases of an odd number of observations and of ties are discussed in the next two sections. Simple devices make the test usable in most cases. A very great tendency toward ties, however, will make it inapplicable: This will be
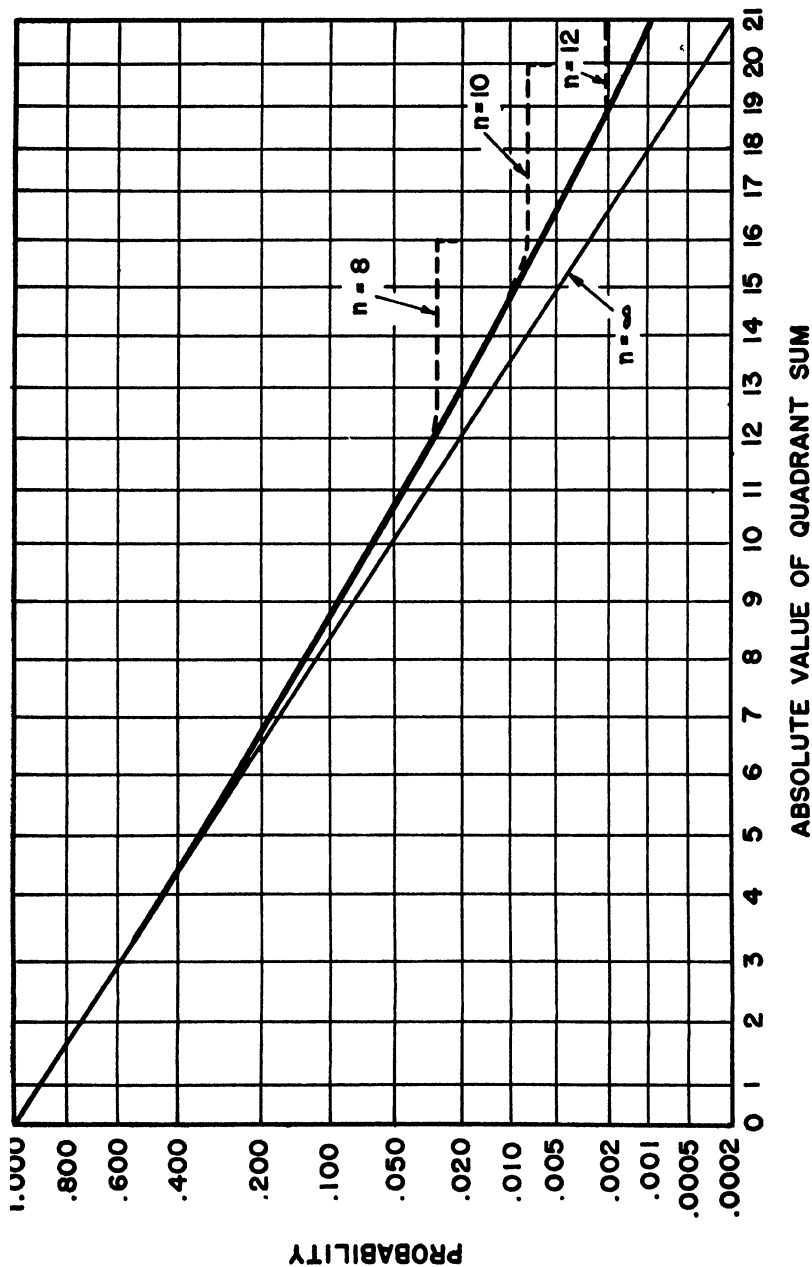
FIG. 1. Probability that an arrangement of a sample of size, N, will have a quadrant sum equal to or greater than indicated absolute value.

unimportant in most applications because of the fact that attention is being directed to the periphery.

INDIVIDUAL TERMS

   TOP = +3
   RIGHT = +1
   BOTTOM = +6
   LEFT = +6 1/2

QUADRANT SUM

   |S| = 16 1/2
   P ≤ 0.5 %
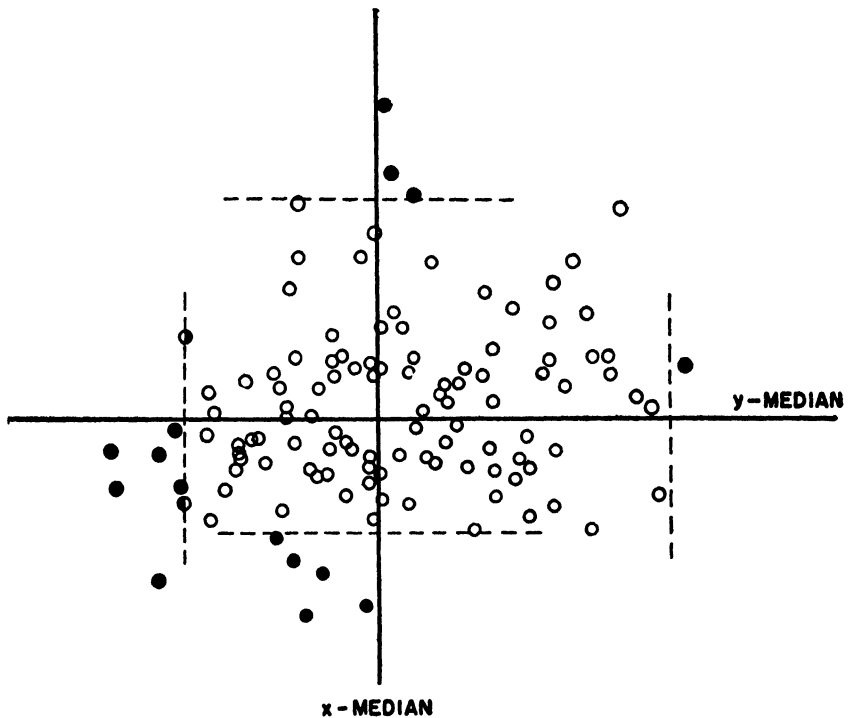

FIG. 2. Scatter diagram of 116 pairs of observations

The set of data which prompted the development of the test is shown in Fig. 2. The accompanying report described it as follows: "The various points appear to be scattered almost completely at random and give little indication of correlation." The quadrant sum is 16½ which is significant at the 0.5% point. Intuitively, the significant association of the peripheral points is clear.

3. Description of test (odd number in sample). If the sample size is odd, then we may usually follow the process outlined above. We will have difficulty only when the counting process meets a point, one of whose coordinates is a median. In this case we employ a simple device, namely:

Given a sample of $2n + 1$ pairs, let x^* and y^* be the medians of the x-values and of the y-values, respectively. Let the pairs in which they occur be (x^*, y_k) and (x_m, y^*), respectively. Replace these two pairs by the sing'e pair (x_m, y_k). There are now $2n$ pairs and the regular method can be applied.

The quadrant sum so obtained from an unassociated population has the same distribution as that formed directly from $2n$ pairs.

4. Description of test (treatment of ties). The behavior of the test is known when (1) there is no association, (2) the probability of a tie in x-values or y-values

TABLE 1

Working significance levels for magnitudes of quadrant sums

Significance level (Conservative)	Magnitude of quadrant sum*
10%	9
5%	11
2%	13
1%	14–15
0.5%	15–17
0.2%	17–19
0.1%	18–21

* The smaller magnitude applies for large sample size, the larger magnitude for small sample size. Magnitudes equal to or greater than twice the sample size less six should not be used.

is zero. The following approximation, which has an unknown effect on the distribution, is suggested when ties are present:

When a tied group is reached, count the number in the tied group favorable to continuing and the number unfavorable. Treat the tied group as if the number of its points preceding the crossing of the median were

$$\frac{\text{number favorable}}{1 + \text{number unfavorable}}.$$

It seems likely that this approximation is conservative.

5. Discussion. When a moderate number, say 25 to 200, of paired observations on two quantities are plotted as a scatter diagram, visual examination frequently detects what seems to be definite evidence of association between the variables. Often in such cases, the usual methods for measuring associa-

tion do not find statistical significance of association. Visual judgment, particularly by engineers or scientists who may wish to take action on the basis of their findings, gives greater weight to observations near the periphery of the scatter diagram. This is not always desirable—but often it is very desirable. A quantitative test of association with such concentration on the periphery has been lacking. The quadrant sum test was developed to fill the gap. Its features of speed and non-parametricity are useful but secondary from this point of view.

When uniform attention to the whole scatter diagram is desired, the quadrant sum test is of unknown usefulness. We know little enough of the operating characteristics of the more conventional tests, such as:

1. The product moment correlation coefficient
2. The four-fold table formed by the medians
3. The biserial correlation coefficient
4. The rank correlation coefficient

and less about the operating characteristics of the present test. In this case, the quadrant sum test can only be recommended definitely for exploratory investigations of large amounts of data.

There are many situations, however, where we do not know where to concentrate our attention, and where speed and non-parametricity are cardinal virtues in a test. One example is the use of serial correlation in studying industrial processes. We may guess that here we are interested in the periphery, but neither theory nor experience can, so far, prove this. In such situations the quadrant sum is by far the fastest to use of any of the tests known to the authors, and we believe one of the most useful.

6. Elementary derivations. We can easily find the distribution of

1. An individual term of the quadrant sum
   a. For fixed sample size
   b. In the limit
2. The quadrant sum itself
   a. For fixed sample size
   b. In the limit, assuming asymptotic independence of the four terms.

This we shall do now, leaving the proof that 2a actually converges to 2b to a later section.

Consider a sample of $2n$ pairs $(x_1, y_1), \cdots, (x_{2n}, y_{2n})$ from a population in which x and y are independent. It is both clear and easily verifiable that

1. The set of $2n$ x-values, $x_1, \cdots, x_{2n}$
2. The set of $2n$ y-values, $y_1, \cdots, y_{2n}$
3. The permutation of the order of the y-values when the pairs are ordered by the x-values

which together determine the sample, are independently distributed, and that any permutation is as likely as every other. (We have assumed no ties, which is a consequence, with probability one, of the continuous cumulative distribu-

tions of x and y). Since the quadrant sum depends only on the permutation, its distribution in the absence of association does not depend on the distributions of x and y.

We must solve, then, certain purely combinatorial problems—under the hypothesis that the $2n!$ permutations of the y-values are all equally likely. It may simplify matters to assume that the values of x in the sample are $1, 2, \cdots,$ $2n$ and that those of y are the same. How, then, do we calculate the distribution of a single term of the quadrant sum. Let us begin with small x-values, and the pair $(1, y_1)$. If $y_1 = 1, 2, \cdots, n$, we count "one" positive, and if $y_1 = n + 1, n + 2, \cdots, 2n$, we count "one" negative. We pass on to $(2, y_2)$ and so on. How many permutations yield a count of exactly k positive values? Those in which $y_1, y_2, \cdots, y_k$ are equal to or less than n, y_{k+1} equal to or greater than $n + 1$, and the other $(2n - k - 1)y$'s are arbitrary. There are:

$$n(n - 1) \cdots (n - k + 1) \cdot (n)(2n - k - 1)!$$

such permutations, the fraction of all $(2n)!$ permutations being:

$$(1) \qquad \frac{n(n - 1) \cdots (n - k + 1)n}{(2n)(2n - 1) \cdots (2n - k + 1)(2n - k)}$$

which is, then, the probability that this contribution will equal $+k$, or by symmetry, the probability that it will equal $-k$, $k \neq 0$.

For large n, this becomes merely:

$$(2) \qquad p_k = 2^{-(|k|+1)}, \qquad k \neq 0.$$

In order to obtain the distribution of the quadrant sum itself, we must concern ourselves with the lack of independence of the four terms. This is indicated most clearly in the case of $2n = 2$, where the $2! = 2$ permutations yield $+1 +1 +1 +1 = 4$ and $-1 -1 -1 -1 = -4$. Here, there is complete lack of independence. We shall see later that there is effectively independence in the limit, so that it is worth while to calculate the sum of four independent terms with the limiting distribution (2) and find that it satisfies:

$$(3) \quad Pr(|\text{ independent sum of 4 terms }| \geq k) = \frac{9k^3 + 9k^2 + 168k + 208}{216 \cdot 2^k}, k > 0.$$

The details will be omitted.

A simple device, reminiscent of Wald's [3, 1943] establishment of the two-dimensional tolerance limits enables us to avoid difficulties with lack of independence and compute the exact distribution of the quadrant sum for any n. We decompose the permutation of the y-values into the following parts, which together specify the permutation:

    (a) The number, j, of pairs in the upper right quadrant.

    (b) The set of j values of x between $n + 1$ and $2n$ corresponding to pairs in the upper right quadrant.

(c) The set of j values of y between $n + 1$ and $2n$ corresponding to points in the upper right quadrant.

(d) The set of j values of x between 1 and n corresponding to pairs in the lower left quadrant. (Note that the use of medians ensures that the lower left and upper right quadrants contain the same number of points.)

(e) The set of j values of y between 1 and n corresponding to pairs in the lower left quadrant.

(f) The permutation of j objects defined by the pairs in the upper right quadrant.

(g) The permutation of $n - j$ objects defined by the pairs in the upper left quadrant.

(h) and (i) the permutations from the remaining quadrants.

It is easily verified that: (1) given j, items (b) to (i) can be assigned at will, (2) each assignment of (a) to (i) corresponds to one and only one permutation, (3) the quadrant sum depends only on items (b) to (e). In fact, the right hand term depends on item (b), the upper term on item (c), the left hand term on item (d) and the lower term on item (e). While j remains fixed, the terms behave independently.

For fixed j, what is the distribution of a single term? If a set of j x-values gives the term $+k$, it must contain the k largest x-values and not contain the next. There are:

$$\binom{n - k - 1}{n - j - 1}$$

such sets. The generating function for a single term, is, then:

(4)
$$\sum_{k=1}^{j} \binom{n - k - 1}{n - j - 1} x^k + \sum_{k=1}^{n-j} \binom{n - k - 1}{j - 1} x^{-k}.$$

Since the terms are independent for fixed j, and there are $(j!)^2((n - j)!)^2$ ways to supply the permutations forming items (f) to (i), the generating function for the quadrant sum, S_n, is:

(5) $$G_n(x) = \sum_{j=0}^{n} \frac{(j!)^2((n - j)!)^2}{(2n)!} \left[\sum_{k=1}^{j} \binom{n - k - 1}{n - j - 1} x^k + \sum_{k=1}^{n-j} \binom{n - k - 1}{j - 1} x^{-k} \right]^4.$$

The exact probability of equalling or exceeding each value of S_n has been computed for $2n = 2, 4, 6, 8, 10,$ and 14. Table 2 gives these probabilities and Fig. 3 shows the values of

$$\frac{m}{5} + \log_{10} Pr(\,|\,\text{quadrant sum}\,|\, \geq m)$$

this particular function being chosen for its relative constancy. The maximum value of the quadrant sum is $4n$, and for values of k less than $4n - 6$, there

TABLE 2

Probability of a Sum of Absolute Value Equal to or Greater than k when a Sample of 2n is Drawn from an Unassociated Population

$|k|$ \ $2n$	2	4	6	8	10	12	14	∞ *
0	1.0000	1.0000	1.0000	1.0000	1.0000		1.0000	1.000000
1	1.0000	0.7500	0.9333	0.9036	0.9106		0.9115	0.912037
2	1.0000	0.7500	0.7556	0.7544	0.7567		0.7580	0.754630
3	1.0000	0.4167	0.6000	0.6000	0.6008		0.6039	0.599537
4	1.0000	0.4167	0.4667	0.4619	0.4662		0.4690	0.462963
5	0.0000	0.3333	0.3111	0.3508	0.3519		0.3547	0.346933
6	0.0000	0.3333	0.2222	0.2619	0.2589		0.2611	0.252025
7	0.0000	0.3333	0.1556	0.1821	0.1867		0.1876	0.177662
8	0.0000	0.3333	0.1111	0.1258	0.1333		0.1322	0.121817
9	0.0000	0.0000	0.1000	0.0839	0.0928		0.0918	0.081471
10	0.0000	0.0000	0.1000	0.0554	0.0642		0.0632	0.053295
11	0.0000	0.0000	0.1000	0.0375	0.0436		0.0432	0.034189
12	0.0000	0.0000	0.1000	0.0304	0.0290		0.0296	0.021557
13	0.0000	0.0000	0.0000	0.0286	0.0190		0.0202	0.013386
14	0.0000	0.0000	0.0000	0.0286	0.0127		0.0139	0.008200
15	0.0000	0.0000	0.0000	0.0286	0.0095		0.0096	0.004963
16	0.0000	0.0000	0.0000	0.0286	0.0083		0.0066	0.002972
17	0.0000	0.0000	0.0000	0.0000	0.0079		0.0045	0.001762
18	0.0000	0.0000	0.0000	0.0000	0.0079		0.0031	0.001036
19	0.0000	0.0000	0.0000	0.0000	0.0079		0.0021	0.000604
20	0.0000	0.0000	0.0000	0.0000	0.0079		0.0014	0.000350
21	0.0000	0.0000	0.0000	0.0000	0.0000		0.0010	0.000201
22	0.0000	0.0000	0.0000	0.0000	0.0000		0.0008	0.000115
23	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000065
24	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000036
25	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000020
26	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000011
27	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000006
28	0.0000	0.0000	0.0000	0.0000	0.0000		0.0006	0.000003
29	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.000002
30	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.000001
31 or over	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000	0.000000
Variance of k	16	24	$26\frac{5}{8}$	$26\frac{4}{7}$	$26\frac{10}{21}$	$26\frac{6}{11}$	$26\frac{140}{143}$	24

' Probability for $2n = \infty$, $k > 0$, is given by

$$\frac{9k^3 + 9k^2 + 168k + 208}{216 \cdot 2^k}.$$

502

is quite good agreement between the curves for finite n and formula (3) at the practically significant percentage points. The situation for very small
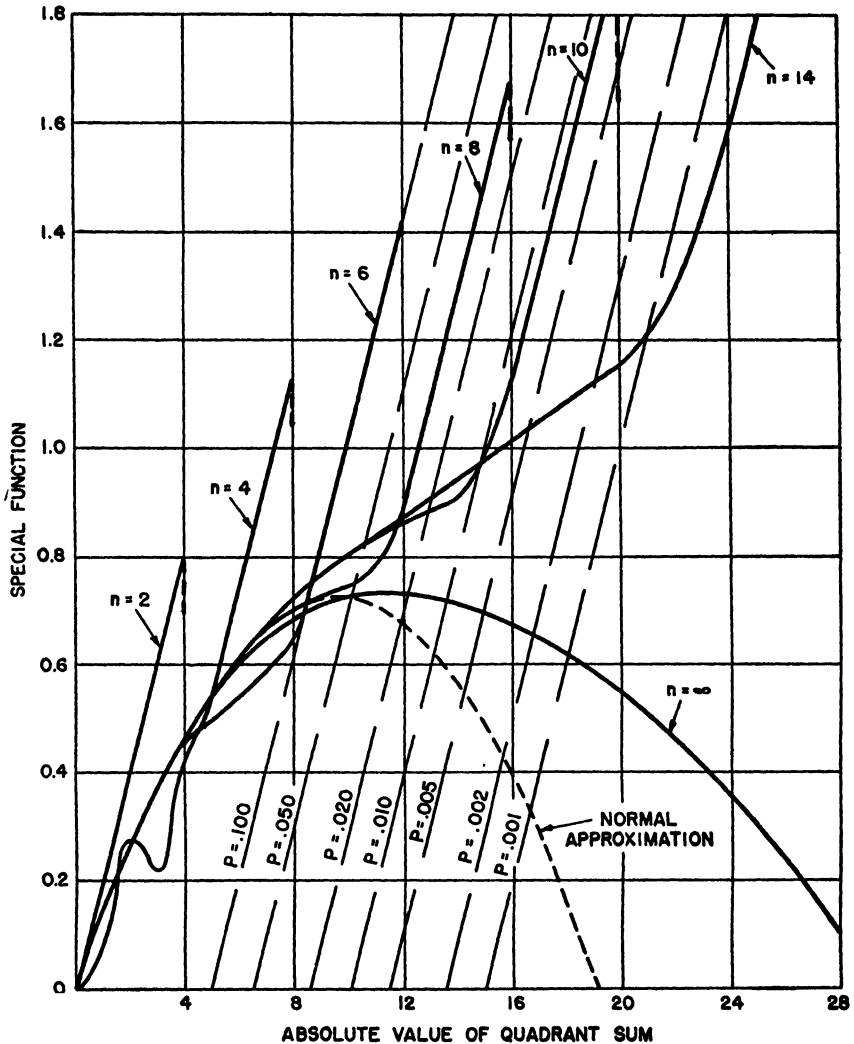

FIG. 3. Comparative relationships for finite and infinite sample sizes and normal approximation to the infinite sample size

probabilities suggests a careful consideration of the limiting behavior of the quadrant sum distribution (see section 10).

The device for samples of $2n + 1$ deserves a word of justification. If there is no association, the $2n + 1$ y-values are randomly paired with the $2n + 1$ x-values, and, in particular, the y-value paired with the x-median is randomly selected. If we pair it with the (randomly selected) x-value which was paired with the y-median we still have random pairing. The pairing of the $2n$ pairs is random, although neither the x-values nor the y-values make up a sample. The randomness of pairing is all that has been used in the discussions of this section.

7. Extension to higher dimensions. The same ideas that underlie the quadrant sum test for two variables may be extended *in several ways* to give tests for various types of association among three or more variables. Only one three-variable case will be discussed here, leaving further extension to the reader.

Given three variables, x, y, and z, and a sample of matched observations on these, it is clearly possible to use the simple quadrant sum test for two variables to investigate association between x and y separately, between y and z separately, and between z and x separately. If the Pearson coefficient of correlation were being computed and were found to be close to zero for each of these pairs, it would be assumed that there was no detectable association through the second moments. In a trivariate normal or Gaussian distribution, where the first and second moments determine the whole distribution, if there is independence between the separate pairs of variables, there is no possibility of a three-way association. It is of some interest, however, to notice that a corner sum test can be devised that will measure the effect of such triple association in case it does exist.

Consider the octants into which the three median planes for x, y, and z, respectively, divide the three dimensional scatter diagram and label the octants alternately plus and minus, in the manner suggested by Fig. 4. More precisely, an octant is counted as plus if an odd number, that is three or one, of the variables are greater than the medians of the sample, and the remaining octants are labelled minus. It is clear that we may repeat the process of coming in along each axis passing from observation to observation as long as they remain in a region of fixed sign, and writing down as a contribution to the final or octant sum the number of such consecutive elements and the sign of the region in which they were found. There will be six terms rather than four, as was the case for the test based on quadrants, and so a new set of significance levels will be required. Table 3, following, lists the situation for a very large sample.

The situation has been sketched for the case of $2n$ triples. If there are $2n + 1$ triples, then we may have trouble with the medians again. However, a similar device works, except that we must agree on a last variable in order to form the synthetic triples uniquely. For example, consider the triples $(m, 3, 5)$, $(9, m, 1)$, $(12, 4, m)$, where m denotes the median. Taking the order in which the variables are written, we get $(12, 3, 5)$ and $(9, 4, 1)$ as the synthetic triples. Other
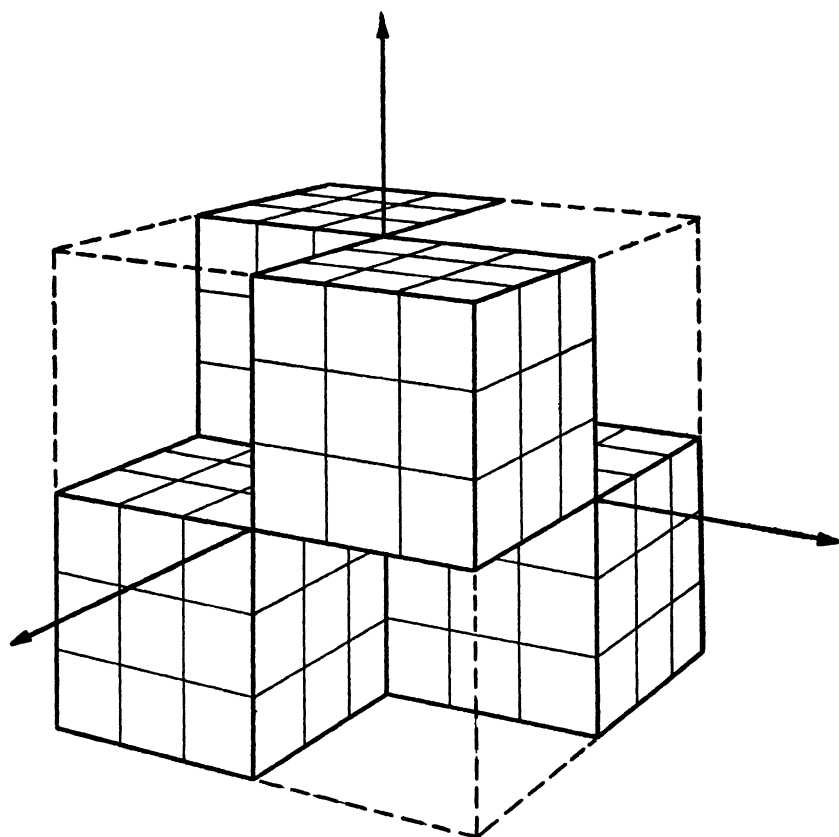
Fig. 4. Octant schematic—solid sections taken as positive

TABLE 3

Working significance levels for the magnitudes of the octant sum

Significance Level	Magnitude of Octant Sum*
10%	11
5%	13
2%	15
1%	16
0.5%	18
0.2%	20
0.1%	21

* Computed for large samples only and based on normal approximation, see section 11 for discussion of this and higher dimensional cases.

orders would yield (9, 3, 5) and (12, 4, 1) or (9, 3, 1) and (12, 4, 5). This slight dissymmetry is not pleasing but should give no difficulty.

8. Nongraphical example. The following example of 78 successive observations of four variables shows how this test may be applied without plotting and how simple the computation still remains. The data concern a metallurgical

TABLE 4

Excerpt from Tippett's Table

Time T^*	Fuel F^*	Material M^*	Articles A^*	Duration D^*
1 −	246 +	1457 −	1895 +	168.5 +
2 −	196 −	2078 +	2121 +	152 +
3 −	192 −	1278 −	1437 −	153 +
4 −	202 +	1398 −	1497 −	145 −
5 −	206 +	1944 +	1592 +	153 +
6 −	218 +	1464 −	1506 −	147.5 −
7 −	155 −	1541 +	1762 +	152 +
8 −	201 +	1502 +	1818 +	144.5 −
9 −	211 +	1950 +	1144 −	151.5 +
10 −	236 +	1768 +	1654 +	151.5 +
etc. to				
78 +	185 −	1536 +	1442 −	152 +
Median 39.5	Median 199	Median 1474	Median 1588	Median 149.5

* Location of observation relative to column median; + = above; − = below.
Tippett's correlations (based on lightly rounded data)

$$r_{FM} = + 0.243$$
$$r_{FA} = + 0.266$$
$$r_{MA} = + 0.681$$
$$r_{FM.A} = + 0.088$$
$$r_{FMA.} = + 0.141.$$

problem in mass production and are taken from L. H. C. Tippett, Table XXII, page 63 [2]. An excerpt from the data is given in Table 4 together with Tippett's calculated correlations. This table also shows the preliminary marking of each individual measurement as above (+) for its variable, below (−), or on the median (0). From this table we see, for example, that increasing T contributes a term −3 to the quadrant sum for T and D. It is often desirable to prepare auxiliary tables to assist in computing the components of the quadrant

and hyperquadrant sums. Such a table is Table 5 for low values of Fuel $(F-)$ arranged in consecutive ascending numerical order. The entries on this table for the five columns headed F, T, M, A, and D are directly comparable to the entries in Table 4. For example, $F = 155$ is $-$ with respect to the fuel median and $T = 7$, $-$; $M = 1541$, $+$; $A = 1762$, $+$; $D = 152$, $+$. The double, triple, quadruple and quintuple headed columns contain simply the algebraic multiplication of the signs in the appropriate T, M, A, or D columns. Thus, TM for $F = 155$ is $-$, MAD is $+$, and $TMAD$ is $-$. The contribution to each quadrant or hyperquadrant sum is simply the count of the consecutive like signs from the top of a column. For column AD, we have 7 consecutive $+$ signs and since the contribution is to FAD and F is $-$, the contribution in this case to the octant sum is -7. The results from the ten tables of which Table 5

TABLE 5

Sample Table for One Component of Quadrant and Hyperquadrant Sums. Low Values of Fuel $(F-)$

Fuel F	T	M	A	D	TM	TA	TD	MA	MD	AD	TMA	TMD	TAD	MAD	TMAD
98 −	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−
135 −	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−
140 −	−	−	−	−	+	+	+	+	+	−	−	−	−	−	+
146 −	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+
147 −	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+
149 −	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−
151 −	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−
153 −	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+
155 −	−	+	+	+	−	−	−	+	+	+	−	−	−	+	−

Contributions to Sums

FT	FM	FA	FD	FTM	FTA	FTD	FMA	FMD	FAD	FTMA	FTMD	FTAD	FMAD	FTMAD
−2	+4	+7	+8	+2	+2	+2	−4	−4	−7	−2	−2	−2	+4	+2

is a sample are then carried to the summary computation shown in Table 6. The contribution from Table 5 is shown on line $F-$. The totals are computed and their probabilities of occurrence determined.

9. Serial example. The following example, a sample of 144 observations of the thickness of inlay for relay springs cut consecutively from a single sheet of material, allows us to compare the resolution of the present test with that of the serial product-moment correlation. The data are from Shewhart [1, 1941, Table 1] and the serial correlations from lag 1 to lag 22 are from recent calculations by Miss Dorothy T. Angell. The procedure for calculating the serial quadrant sums is similar to that for obtaining the sums for section 8. A table is prepared to show the observed consecutive order of the numerical values and each is identified as above $(+)$, below $(-)$, or on the median (0). This gives a

TABLE 6

Summary Computation Table for Quadrant and Hyperquadrant Sums

From Table	TF	TM	TA	TD	PM	PA	PD	MA	MD	AD	TPM	TPA	TPD	TMA	TMD	TAD	PMA	PMD	PAD	MAD	TPMA	TPMD	TPAD	TMAD	PMAD	PMAD	TPMAD
$T+$	$-6+1$	$-2+3$	$-2+1$			$+1+1$					$-1+2$	$+2-3$		$-1-1$	$+1+1$	-2			$+3$		$+1-1$	$-1-1$	$+1+2$				$+1+4$
$T-$	$-1+1$	$-2-3$				$+7+8$					$+1+1$	$-1-1$	$+1+1$	$+1$		-2			-7		$+3-2$	$+1+1$	$-1-1$				$+2+4$
$F+$	$+4$				$+4+4$		$+1+8$	$+5+25$	$+13+3$		$+5+2$	$+1+1$	$+2+2$	$+1+2$	$+1+2$		$+1+1$	$+2+3$		$+5-3$	$+1-2$	$+3-2$	$+3-2$		$+2+3$	$+3+4$	$+3+2$
$F-$	-2	-2			$+2+6$	$-1+5$	$+2+3$	$+3+1$	$+3$		$+2$			$+2$			-4	$-6-3$	-7	$-3-1$	$+1-2$		$-1-2$		$+2+3$	$+2+3$	$+1+2$
$M+$		$+1$		$+1$						$+7+4$	$+1+2$		$+1+1$		$+1+2$	$-2+1$	$-2-1$		$-1-4$	$+1+3$	$+1-1$		$+1-2$		-1	$+1$	$+1+3$
$M-$		-2		-1							$+2$		$+1$		$+2$		-1		$+2-3$	$-1-1$	-2		-3		$+1$	$+2+3$	$+2+3$
$A+$								$+3+1$	$+1+1$	$+2+7$			$+1+1$		$+3+2$	$+3+1$	$+1-1$	$+1$	$+2-3$	$-2+1$		$+1-3$	$-2+1$		$-1+1$	$-5+1$	$+1+3$
$A-$													$+1+1$			$+2+1$				$+3+1$		$+2$	$+2$		$+1$	$+14$	$+2+3$
$D+$																				$-10+4$							
$D-$																											
Totals Quadrant Sums	-5	$+1$	-7	0	$+16$	$+12$	$+14$	$+34$	$+18$	$+20$																	
Octant Sums																											
Hexadecant Sums											$+11$	$+6$	$+1$	$+4$	$+10$	-3	-9	-4	-10	$+4$	0	-1	$+1$	-5	$+14$		
Dotriacontant Sums																											$+22$
Probability (%) $\leq$	36	92	19	100	0.2	2	1	0.1	0.1	0.1	9	37	94	57	12	68	16	57	12	57	100	95	95	52	6		0.6
Significant at 5%					$*$	$*$	$*$	$*$	$*$	$*$																	$*$
Significant at 1%					$*$			$*$	$*$	$*$																	$*$
Significant at 0.2%					$*$			$*$	$*$	$*$																	

table similar to one of the elements, say Fuel, in Table 4. Four computation tables similar to Table 5 are required, one for the equivalent of moving from the right, one from below, one from the left, and one from the top of a lag correlation scatter diagram. One table from each direction will take care of all lags. In the first, the marginal entries are the observed values listed in descending numerical order. Opposite these are recorded from the previous table the signs associated with observations for each lag with respect to each entry. The second table would record the signs relating to the lags from the observed values arranged in ascending order. The third table would record the signs relating to leads from the observed values arranged in ascending order and the fourth, the signs relating to leads from the observed values arranged in descending order. The sign of the contribution from each group is the algebraic product of the sign of the run and the sign of the marginal entries. The length of run is determined in the same way as in Table 5. Table 7 illustrates the procedure

TABLE 7

Relation of Lagged Observations to Median (+ = above, − = below) for Smallest Observations in Ascending Order

Thick-ness	Lag																									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	−	−	+	−	−	+	−	+	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−
8	−	−	+	−	+	−	−	−	−	+	+	−	+	+	−	+	−	+	+	+	+	+	+	+	+	+
8	−	−	−	+	−	−	−	−	+	+	+	−	+	+	−	+	+	−	+	+	−	−	−	−	−	−
10	−	+	−	−	−	−	+	+	−	+	+	−	+	+	−	−	+	+	−	−	−	−	−	−	+	−
13	−	−	−	−	−	+	+	+	−	+	+	+	+	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
17	−	−	−	+	−	−	+	−	+	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−
17	−	+	−	+	+	+	+	+	+	−	−	+	+	−	−	−	−	−	+	−	−	+	−	+	+	
18	−	−	−	−	+	+	−	+	+	−	+	+	−	−	+	+	−	−	−	−	−	+	−	+	−	
•		+3	−2	+2	+1	−1	+3	−1	+2	−5	+3	−1	−7	−1	−3	−1	−2	−1	−3	−3	−2	−2	−2	−2	+1	+1

• Contribution to Serial Quadrant Sum.

of determining the contribution from lags associated with the observations arranged in ascending order.

Two serial quadrant sums may be computed—a circular serial quadrant sum or a noncircular serial quadrant sum. Circular items arise from considering that the beginning of the set of observations is a continuation of the end in the same way that this assumption is made in computing circular serial correlation coefficients. In Table 7, circular items are shown in parentheses and are omitted in calculating noncircular sums. In the particular table shown, the count of the run lengths was identical for both types of sum, but in other cases this may not be the case. Since the serial quadrant sum is relatively insensitive to sample size, the noncircular serial quadrant sum has for all practical purposes the same distribution as the circular quadrant sum. The correspondence in this case between the serial correlation coefficient for each lag up to 22 and the respective values of the two types of serial quadrant sums is shown in Fig. 5.
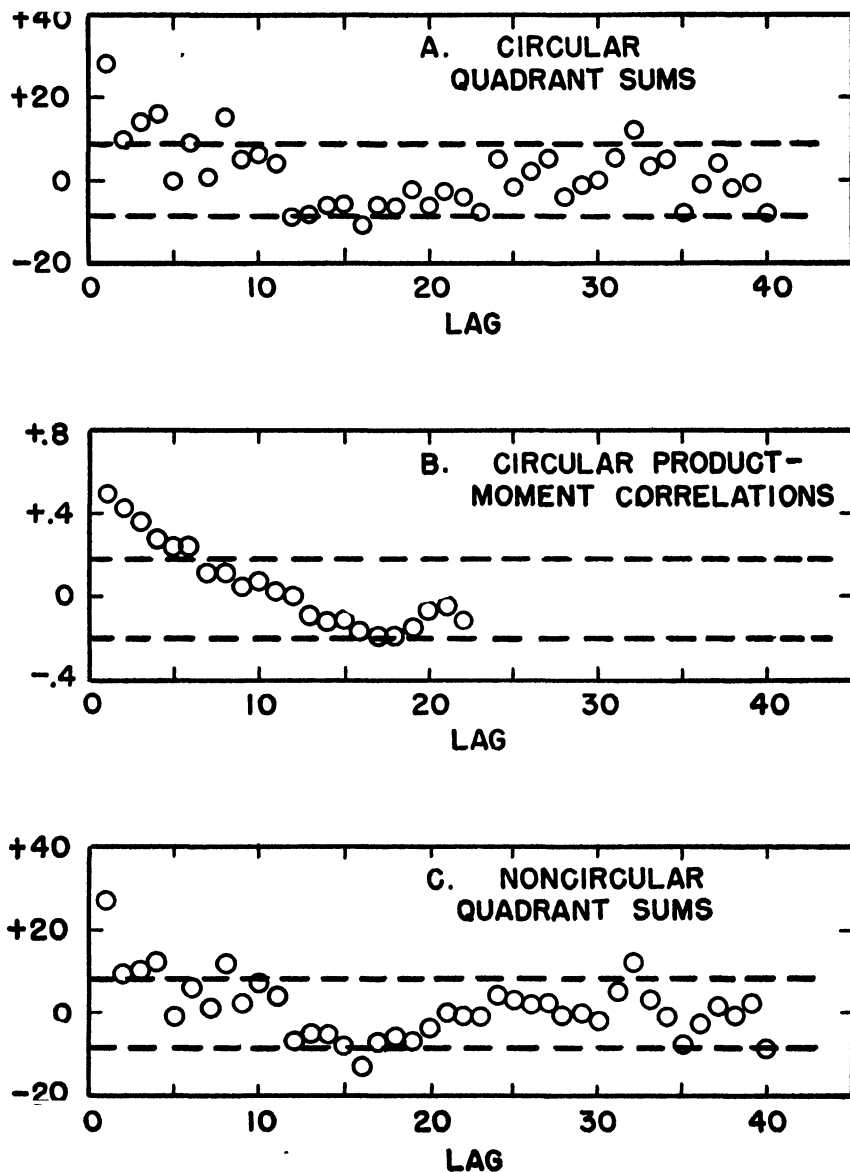
FIG. 5. Comparative performance on a serial (autocorrelative) example

10. Convergence to the limiting distribution. We shall consider several chance sums. One of these is S, which has the limiting distribution discussed

in section 6. Another is S'_k, which is the sum of four independent terms, each distributed according to the limiting distribution curtailed at $\pm k$. Its generating function is

$$G_k(x) = \left(\sum_{i=1}^{k} 2^{-(i+1)} x^i + \sum_{i=1}^{k} 2^{-(i+1)} x^{-i} \right)^4.$$

The total probability assigned to $S'_k = -k, -(k-1), \cdots, k$, is less than unity, so that there is nonzero probability that S'_k is not defined. The third is S_n, the quadrant sum itself, whose generating function is (5), and the fourth is the result of the same sort of curtailment applied to S_n. It will be denoted by $S_{n,k}$ and its generating function is

$$G_{n,k}(x) = \sum_j \frac{(j!)^2((n-j)!)^2}{(2n)!} \left(\sum_{i=1}^{k} \binom{n-i-1}{n-j-1} x^i + \sum_{i=1}^{k} \binom{n-i-1}{j-1} x^{-i} \right)^4.$$

This again corresponds to a total probability less than unity.

It is clear that

$$\Pr(S_{n,k} = m) \leq \Pr(S_n = m)$$

and

$$\Pr(S'_k = m) \leq \Pr(S = m).$$

We shall soon show that

(6) $$\lim_{n \to \infty} \Pr(S_{n,k} = m) = \Pr(S'_k = m)$$

and this will imply that

$$\lim_{n \to \infty} \Pr(S_n = m) = \Pr(S = m)$$

which is the desired result. The implication runs as follows: given ϵ, we can choose k so large that

$$\Pr(S'_k \text{ defined}] \geq 1 - \epsilon/3$$

whence

$$| \Pr(S'_k = m) - \Pr(S = m) | \leq \epsilon/3$$

and then choose n so large that

$$| \Pr(S_{n,k} = m) - \Pr(S_k = m) | \leq \epsilon/(24k + 6)$$

$$\text{for } m = -4k, -4k + 1, \cdots, 4k$$

whence

$$\Pr(S_{n,k} \text{ defined}) \geq 1 - \epsilon/3 - \frac{8k+1}{24k+6} \epsilon \leq 1 - \frac{16k+3}{24k+6} \epsilon$$

and hence

$$| \Pr(S_{n,k} = m) - \Pr(S_n = m) | \leq \frac{16k + 2}{24k + 6} \epsilon$$

this inequality holding automatically for $| m | > 4k$.   Hence,

$$| \Pr(S_n = m) - \Pr(S = m) |$$
$$\leq | \Pr(S_n = m) - \Pr(S_{n,k} = m) | + | \Pr(S_{n,k} = m) - \Pr(S'_k = m) |$$
$$+ | \Pr(S'_k = m) - \Pr(S = m) | \leq \frac{16k + 2}{24k + 6} \epsilon + \frac{1}{24k + 6} \epsilon + \tfrac{1}{3}\epsilon < \epsilon$$

This method is clearly of general application in such problems.

We turn now to the proof of (6).   The expression for $G_{n,k}(x)$ shows that we may consider it the result of the following process: the integer j is a chance quantity with the distribution

$$\Pr(j = j_0) = \frac{(n!)^2}{(2n)!} \binom{n}{j_0}^2 .$$

For fixed j, $G_{n,k}$ is the average over j of

$$G_{n,k,j}(x) = \left[\sum_{i=1}^{k} \frac{\binom{n - i - 1}{n - j - 1}}{\binom{n}{j}} x^i + \sum_{i=1}^{k} \frac{\binom{n - i - 1}{j - 1}}{\binom{n}{j}} x^{-i} \right]^4 .$$

The first of these relations shows that j/n converges stochastically to $\tfrac{1}{2}$ as n approaches infinity.   The second shows, since

$$\frac{\binom{n - i - 1}{n - j - 1}}{\binom{n}{j}} = \frac{(n - i - 1)!(n - j)!j!}{(n - j - 1)!(j - i)!n!} = \frac{(n - j)(j)(j - 1) \cdots (j - i + 1)}{n(n - 1)(n - 2) \cdots (n - i)}$$

$$\frac{\binom{n - i - 1}{j - 1}}{\binom{n}{j}} = \frac{(n - i - 1)!(n - j)!j!}{(n - j - i)!(j - 1)!n!}$$
$$= \frac{(n - j)(n - j - 1) \cdots (n - j - i + 1)j}{n(n - 1) \cdots (n - i)}$$

and both of these converge stochastically to $2^{-(i+1)}$ as n approaches infinity, that $G_{n,k,j}(x)$ converges stochastically to $G_k(x)$.   Since these curtailed generating functions involve only powers of x in the finite range between $-4k$ and $+4k$, the limiting relation (6) follows at once.

11. Effectiveness of normal approximation.   Fig. 3 shows the relation between the asymptotic distribution of the quadrant sum for large n and a normal

distribution with variance 24, i.e., the same variance as that of the asymptotic distribution. The normal approximation is calculated from

$$\Pr(|\,S_n\,| \geq m) \approx \Pr\left(x \geq \frac{m - \frac{1}{2}}{\sqrt{24}}\right)$$

where x is normally distributed with zero mean and unit variance. The asymptotic and normal curves agree surprisingly well out to the 5% point, and an error of a full unit in the significance level first occurs beyond the 0.5% point.

Since the asymptotic distributions for the quadrant, octant, hexadecant, dotriacontant,—, sums become more and more normal, the normal approximation will be even better for higher dimensions. In r dimensions, this approximation consists in treating

$$\frac{|\,S_n\,| - \frac{1}{2}}{\sqrt{12r}}$$

as the absolute value of a standard deviate. This should be quite adequate for large samples and $r \geq 4$.

12. Unsolved problems. The central unsolved problem in connection with the quadrant sum is:

(1) What is the operating characteristic?
This has as a corollary the more general question:
  (2) How can the operating characteristic of a nonparametric test be described so as to be useful to the users of the test?
There are, of course, minor problems which are much more easily soluble. A few, listed in order of practical importance, are:
  (3) What is the effect on the significance levels of the use of lagged values of x as values of y?
  (4) What are the exact distributions for moderate n in three or more dimensions?
  (5) Do the analogous limiting distributions hold for three or more dimensions?
  (6) What is a better approximation to the limiting distribution for moderate n?

To encourage others to solve some of these, we close with the assurance that they have our good wishes.

REFERENCES

[1] W. A. SHEWHART, "Contribution of statistics to the science of engineering," *University of Pennsylvania Bicentennial Conference*, Volume on Fluid Mechanics and Statistical Methods in Engineering, University of Penn. Press, 1941, pp. 97-124. Also Monograph B-1319, *Bell Tel. System Tech. Pub.*, 1941.
[2] L. H. C. TIPPETT, *Statistical Methods in Industry*, pamphlet published by Iron and Steel Research Council of the British Iron and Steel Federation, 1943.
[3] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 45-55.

DISCRIMINANT FUNCTIONS

By George W. Brown

Iowa State College

1. Introduction: In the following sections the development of discriminant function techniques is approached from an elementary point of view, considering first an essentially trivial problem, then working up to the more complex situations which may be handled by discriminant function methods. No attempt has been made to follow the pattern of the historical development in this process, and no consistent attempt has been made to allocate proper credit, in the text, to those individuals responsible for the introduction and exploitation of these methods. A more or less exhaustive bibliography of discriminant function applications and related theory is given at the end of this paper.

Some historical perspective may be gained, however, from a very sketchy consideration of the early background of the subject. The first published application of the discriminant function seems to have been the work of Barnard (1935 [1]) on craniometry, following the suggestion of R. A. Fisher. Meanwhile P. C. Mahalanobis (1927, [30]; 1930, [31]) and, in this country, Hotelling (1931, [25]) had been concerned with a closely related problem, the construction of measures of the "distance" between two sets of multiple measurements, for which Karl Pearson's (1926, [34]) coefficient of racial likeness was not wholly adequate. Fisher (1936, [18]) gave a further example of the method and showed (1938, [19]) the relation between his work and that of Hotelling (1931, [25]; 1936, [27]). Thus the theory of discriminant function analysis proper is about ten years old, but is intimately related to researches which go back a few more years.

A simple problem: Consider the very simple case of a single measurement, say ξ, which may be made in each of two populations, and let us suppose, for the sake of discussion, that ξ is normally distributed, with unit variance, in each population, but with possibly different means in the two populations.

Let

$$E_1(\xi) = \alpha - \beta$$

$$E_2(\xi) = \alpha + \beta$$

be the mean values of ξ over the two populations, with $\beta > 0$. As an example, we may consider the pH measurements of Iowa soil samples (Cox and Martin, [12]), for two soil populations, distinguished by the presence or absence of Azotobacter. From 100 samples containing Azotobacter and 186 samples containing no Azotobacter, we have the estimated averages of pH equal to 7.423 and 6.015 respectively, with an estimated standard error of .625 within populations (see Fig. 1).

$$\hat{\alpha} = 6.719$$

$$\hat{\beta} = .704$$

$$\hat{\sigma} = .625$$

$$\hat{\beta}/\hat{\sigma} = 1.13.$$

Let us suppose further that ξ is the only measurement available on a single individual, not knowing to which of populations 1 and 2 the individual belongs.

Distribution of pH Measurements

With Azotobacter   Without Azotobacter

6.015   6.719   7.423
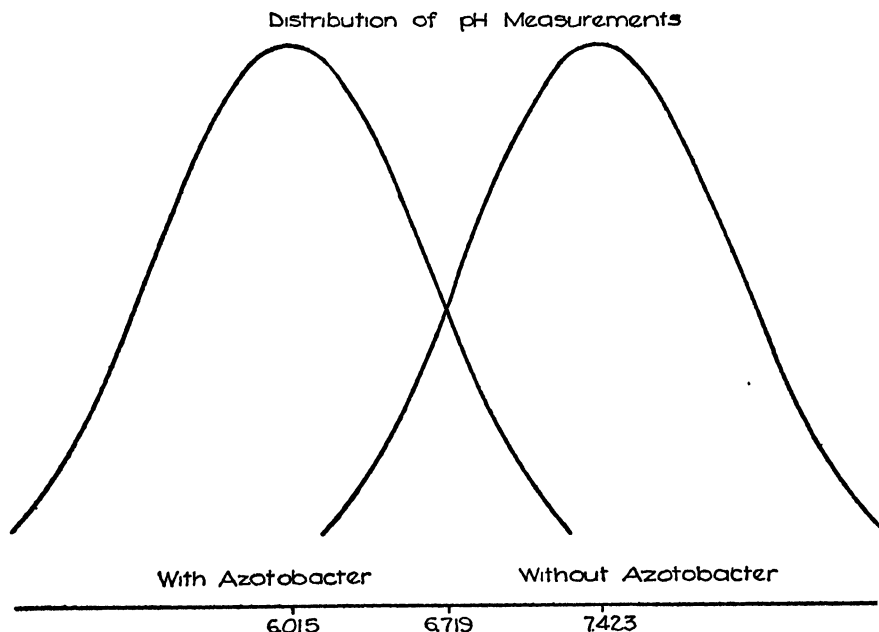
FIG. 1

The problem is to classify this individual as a member of population 1 or population 2. It is clear that ξ furnishes the only information on which to base a decision, and that essentially the only procedure available is to choose a number, say ξ_0, such that we choose population 1 when $\xi < \xi_0$ and population 2 when $\xi > \xi_0$. Furthermore, it is evident that the expected accuracy of classification depends on the size of β. If we wish to have equal risks of misclassification for members of the two populations we choose $\xi_0 = \alpha$. Then the probability of misclassification is given by $P\{\epsilon > \beta\}$, where ϵ is a normal deviate with unit variance. As one would expect, the probability of misclassification tends to 0 as $\beta \to \infty$ and tends to $\frac{1}{2}$ as $\beta \to 0$. In the Azotobacter example, if we assume that the estimates given are the population values, we choose $\xi_0 = 6.719$. The

ratio $\hat{\beta}/\dot{\sigma} = 1.13$ is exceeded approximately 13% of the time in sampling from the normal distribution, leading to .13 as the probability of misclassification.

Consider now the slightly more general situation in which we consider a fixed variate, say w with measurements ξ distributed, for fixed w, with a mean of the form $\alpha + \beta w$. This is the standard regression situation. As before assume that ξ is normally distributed about this mean with unit variance, that is

$$\xi = \alpha + \beta w + \epsilon$$

where α and β are constants, w may take on any or all real values, and ϵ is a normal deviate. Note that if w is restricted to take on only two values the structure reduces to the first structure considered. An example of the continuous type might be constructed by considering w as genotypic yield of grain and ξ a phenotypic measure of yield (Smith, [36]).

The simple problem formulated for the two-population case may be reformulated here as follows: Given the relationship $\xi = \alpha + \beta w + \epsilon$, and given ξ for an individual for which no other information is known, how shall we estimate w? For selective breeding the problem may be to select individuals for which w is at one end of the scale, rather than to estimate w itself. Whatever decision is to be made, it is still clear that ξ furnishes the only available information, and that the certainty of the decision is a function of β. Since $(\xi - \alpha)/\beta = w + \epsilon/\beta$, the variance of this estimate of w is $1/\beta^2$. Note that confidence intervals for w, given ξ, may be constructed from the normally distributed quantity $\xi - \alpha - \beta w$.

It should be pointed out that in the usual regression case we are interested in predicting ξ for given w, with the hypothesis as stated above, whereas in this case ξ will be observed, and the problem is that of estimating, as a parameter of the distribution of ξ, the fixed variate w.

Obviously β must not vanish if ξ is to perform any discrimination among w values. In practice, of course, α and β will not be given as known values and the variance of ϵ will not be known, but a finite set of observations may be available, for which w values are known and ξ has been observed. The usual analysis of variance provides a significance test for the non-vanishing of β, which is equivalent to testing for the significance of the regression of ξ on w.

It is to be noted that this analysis reduces to the conventional between-within analysis (F or t-test) when we have the special case of two populations. Moreover, if we had treated ξ as the fixed variate instead of w, and considered the regression of w on ξ, the Analysis of Variance would have differed only in replacing $\Sigma(\xi - \bar{\xi})^2$ throughout by $\Sigma(w - \bar{w})^2$ and the relevant F-test would have been unchanged.

When probabilities of misclassification are estimated from finite samples, as in the soil classification example, there are three sources of error, sampling error in the estimate of the separation value ξ_0, sampling error in the estimate of the distance between the population means, and sampling error in the estimated standard deviation of ξ within populations. It does not appear difficult to set up confidence intervals for the probability of misclassification, assuming repeated classification of individuals given fixed initial samples.

2. The one-dimensional discriminant function. We have been dealing so far with the simple situation in which only one measurement per individual is available for purposes of discrimination. Suppose we still have this measurement, call it ξ_1, now, but we have other measurements as well, say $\xi_2, \cdots, \xi_p$. As before $\xi_1 = \alpha_1 + \beta w + \epsilon_1$. For the moment suppose that the remaining measurements have mean values independent of w, so that

$$\xi_m = \alpha_m + \epsilon_m, \qquad (m = 2, \cdots, p),$$

and let us assume also that the $\{\epsilon_m\}$ are mutually independent, $(m = 1, 2, \cdots, p)$ and are normal deviates with unit variance. It is safe to assume that nobody would ever argue, in this case, that the measurements $\xi_2, \cdots, \xi_p$, provide information about the w value for an individual. If, then, we were so fortunate that we were in this situation, and knew so, we could say that ξ_1 is our discriminant function, since, if any discriminating is to be done, ξ_1 has to do it.

TABLE 1

Analysis of Variance for Regression

	d.f.	Sums of Squares
Regression	1	$r^2 \Sigma(\xi - \bar{\xi})^2$
Error	$N - 2$	$(1 - r^2)\Sigma(\xi - \bar{\xi})^2$
Total	$N - 1$	$\Sigma(\xi - \bar{\xi})^2$

$$r = \frac{\Sigma(\xi - \bar{\xi})(w - \bar{w})}{\sqrt{\Sigma(\xi - \bar{\xi})^2 \Sigma(w - \bar{w})^2}}$$

Suppose, now that the measurements $\xi_1, \xi_2, \cdots, \xi_p$ are not explicitly available, but that we are able to observe a linearly equivalent set $x_1, x_2, \cdots, x_p$, related to the $\{\xi_m\}$ by the transformation

$$x_m = \sum_{n=1}^{p} l_{mn} \xi_n$$

where the l_{mn} are unknown. For fixed w, x_m has expected value

$$\sum_{n=1}^{p} l_{mn}\alpha_n + l_{m1}\beta w = a_m + b_m w,$$

so that in general each x_m observation provides information about w. Moreover, the x_m are not in general mutually independent; it is evident that the population matrix of variances and covariances for fixed w is given by $\sigma_{mn} = \sum_{k=1}^{p} l_{mk} l_{nk}$.

As an example of a set of correlated measurements, consider the Azotobacter example referred to above. In addition to pH values, determinations of avail-

able phosphate content and total nitrogen content were made on soil samples
in each of the two populations.   Means were as follows:

	pH	Phosphate	Nitrogen
Mean of 100 samples with Azotobacter	7.423	133.120	29.400
Mean of 186 samples without "	6.015	51.113	21.140
Mean difference	1.408	82.007	8.260

Clearly the differences are proportional to the hypothetical b_m's.   The variance-
covariance matrix, estimated from the 284 degrees of freedom within populations,
is given by Table 2.

<div align="center">TABLE 2</div>

	pH	Phosphate	Nitrogen
pH	111.0879	2,292.7192	198.4026
$284(\sigma_{mn})$ = Phosphate		1,042,799.1890	5,066.2645
Nitrogen			29,422.3655

Estimated correlation coefficients within populations are not large, .213 for pH
and Phosphate, .110 for pH and Nitrogen, and .029 for Phosphate and Nitrogen.

Another example is furnished by Fisher's Iris measurements [8], provid-
ing sepal length, sepal width, petal length, and petal width for each of 50
individuals of Iris setosa and 50 individuals of Iris versicolor.   This example is
an unfortunate one in that either petal length or petal width alone is sufficient
to discriminate the two populations as completely as anybody has a right to
expect anytime.   The petal lengths, for example, vary between 1.0 and 1.9 cm.
for the 50 setosa, and between 3.0 and 5.1 cm. for the 50 versicolor.

Let us proceed, under the assumption that available measurements, x_m,
are distributed normally about mean values $a_m + b_m w$, with variance covari-
ance matrix σ_{mn} for fixed w, keeping in mind the underlying model of ξ_1, ξ_2, $\cdots$,
ξ_p, with

$$x_m = \sum_{n=1}^{p} l_{mn} \xi_n, \qquad \xi_1 = \alpha_1 + \beta w + \epsilon_1; \qquad \xi_2 = \alpha_2 + \epsilon_2; \cdots; \xi_p = \alpha_p + \epsilon_p.$$

The skeptic may wish to grant the first part of our assumptions without grant-
ing the hypothetical structure of ξ's underlying the x's.   Hotelling's work [27]
shows that such an underlying structure of ξ's may always be provided, given
the distribution of x's for fixed w.   In other words, a distribution of x's for fixed
w leads essentially uniquely to an underlying ξ model.

The discriminant function, given σ_{mn}, a_m and b_m, for $m, n, = 1, 2, \cdots, p$,
is

$$X = \sum_{m,n=1}^{p} \sigma^{mn} b_m x_n = \sum_{n=1}^{p} t_n x_n$$

where

$$t_n = \sum_{m=1}^{p} \sigma^{mn} b_m, \quad \text{and} \quad \sigma^{mn}$$

is the reciprocal matrix to σ_{mn}. That is σ^{mn} are the solutions of the linear systems [17]

$$\sum_{s=1}^{p} \sigma^{ms} \sigma_{sn} = 0 \quad \text{if} \quad m = n; \qquad m, n, = 1, 2, \cdots, p$$

$$\sum_{s=1}^{p} \sigma^{ms} \sigma_{sm} = 1; \qquad m = 1, \cdots, p.$$

That X, as defined above, is properly called the discriminant function will become evident immediately. Putting $b_m = l_{m1}\beta$, $x_n = \sum_{k=1}^{p} l_{nk}\xi_k$, we have

$$X = \beta \sum_{m,n,k,} \sigma^{mn} l_{m1} l_{nk} \xi_k.$$

Recalling that the σ^{mn} are reciprocal to $\sigma_{mn} = \sum_k l_{mk} l_{nk}$, it can be seen that $\sum_{mn} \sigma^{mn} l_{m1} l_{nk} = 1$ if $k = 1$, and vanishes for $k \neq 1$. It follows that

$$X = \beta \xi_1,$$

in other words, X calculated as $\sum_{mn} \sigma^{mn} b_m x_n$ from known population quantities is proportional to the hypothetical ξ_1, the only one of the underlying measurements which is related to w, thus justifying the term discriminant function for X. It is clear that any other linear function of the x's is also a linear function of the ξ's, and can discriminate, at best, only as well as X itself, since all the ξ's are independent of w, with the exception of ξ_1. X itself discriminates w to the same extent that ξ_1, were it available, would discriminate.

The degree of discrimination of w's depends, as indicated in the previous section, on the ratio of the mean square of ξ_1, among w's (mean square for regression), to the mean square of ξ_1 for fixed w (mean square for error). Since X is proportional to ξ_1, the same is true when X is substituted for ξ_1. It turns out, of course, that X is that linear combination of x's for which the ratio of the mean square for regression to the mean square for error is a maximum, or, what is the same thing, X is that linear combination of X's which has the maximum correlation with w. From any point of view X appears to be the logical function of x's to compute. It is clear that λX is precisely as good as X, if λ is any constant.

In the two population case, where w takes on only two values, X is evidently proportional to $\Sigma\sigma^{mn}(\mu_{m1} - \mu_{m2})x_n$, where μ_{m1} and μ_{m2} are the mean values of x_m in the two populations. X is here the particular linear combination of x's for which the ratio of the mean square between populations to the mean square within populations is a maximum. The value of this ratio, which measures the degree of discrimination possible, depends on the spread of the means of X between the populations, or in general, on the spread of the means of X over some given distribution of w's. Given σ_{mn} and b_m the larger the spread of w values the better overall discrimination will be obtainable. On the other hand, the coefficients for X depend only on σ_{mn} and b_m.

Since X is proportional to ξ_1, it follows that the discriminant function is invariant under non-singular linear transformation of the x's, that is, if some set of y's, linearly dependent on the x's, had been observed, together with their means, variances and covariances, the discriminant values would not have changed. This invariance is obviously a desirable property, and as such was one of the goals of Fisher, Hotelling, and Mahalanobis. One more property of the discriminant function is of interest; X is essentially equivalent to the maximum likelihood estimate of w.

In our statistical model w plays the role of a fixed variate or population parameter, and the x's have a joint distribution about linear functions of w as means. Suppose now that (σ_{mn}) and $\{b_m\}$ are estimated from an analysis of variance and covariance on data for which w as well as x values are known. The problem of estimating w for a single individual whose x measurements are given resolves into a two-stage estimation process, the first stage being the estimation of (σ_{mn}) and $\{b_m\}$ from the initial data, the second stage being the estimation of w by the discriminant function whose coefficients are computed from the estimated (σ_{mn}) and $\{b_m\}$. It has already been pointed out that X is the linear combination of x's which has greatest correlation with w. It turns out, then, that the coefficients of X are proportional to those which would have been obtained from a formal regression analysis of w on $x_1, x_2, \cdots, x_p$, considering the x's as independent variables and w as dependent variable, a direct interchange of roles as compared with the statistical model we have assumed. Of course two linear functions differing only by a factor of proportionality are equivalent in discrimination. If the formal analysis of variance is carried out for testing the significance of the regression of w on $x_1, x_2, \cdots, x_p$, the relevant F ratio remains a valid test for the non-vanishing of the b_m in spite of the inversion of dependent and independent variables. The analysis of variance is given in Table 3.

R is, of course, the conventional multiple correlation coefficient. An equivalent analysis can be carried out for X itself, allowing sufficient degrees of freedom for the estimation of the constants in X, as given in Table 4.

This analysis is proportional to the analysis given above. It might be noted that the mean square corresponding to error sum of squares in this analysis is $\Sigma\sigma^{mn}b_m b_n$, which is X evaluated for $x_n = b_n, (n = 1, 2, \cdots, p)$.

In the Azotobacter example, Cox and Martin arrive at a discriminant function which has the analysis given in Table 5.

It is evident that the difference between populations is highly significant. The choice of scale for X in this case forces the sum of squares within populations to be equal to the difference between the mean X values for the two populations. Thus the mean X differs by .021777 for the two populations, and has an esti-

TABLE 3

Analysis of Variance for Regression

	d.f.	Sums of Squares
Regression	p	$R^2\Sigma(w - \bar{w})^2$
Error	$N - p - 1$	$(1 - R^2)\Sigma(w - \bar{w})^2$
Total	$N - 1$	$\Sigma(w - \bar{w})^2$

TABLE 4

Analysis of Variance for X on w

	d.f.	Sums of Squares
Regression	p	$R^2\Sigma(X - \bar{X})^2$
Error	$N - p - 1$	$(1 - R^2)\Sigma(X - \bar{X})^2$
Total	$N - 1$	$\Sigma(X - \bar{X})^2$

TABLE 5

Analysis of Variance of Discriminant Function

	d.f.	Sums of Squares	Mean Square
Between populations	3	.030842	.01028
Within populations	282	.021777	.00007722
Total	285		

mated standard error, within populations, equal to $\sqrt{.00007722} = .008788$. Half the difference, divided by the standard error is the normal deviate corresponding to misclassification, if equal risks are taken. In this case the value of the normal deviate is 1.24, approximately, leading to an estimated probability of misclassification of about .11, which is not very much better than the .13 which one would have obtained if pH alone had been used.

In this problem, as in conventional regression analysis, it is tempting, for

various reasons, to consider the possibility of using smaller sets of classifying measurements. Moreover, a significance test for this situation is in general more interesting, as a practical matter, than the significance test for differences among populations, since the initial presumption is that we are interested in being able to discriminate, on the basis of x_1, x_2, $\cdots$, x_p. Suppose, for example, we wish to test whether the discriminant function $X_{(p)}$ based on x_1, x_2, $\cdots$, x_p is significantly better than the discriminant function $X_{(r)}$ based on x_1, $\cdots$, x_r, with $r < p$. The relevant test is precisely the same as the test

TABLE 6

Analysis of Variance for Rejecting x_{r+1} , $\cdots$, x_p

		Sums of Squares	d.f.
S_r^2	Regression on	x_1 , $\cdots$, x_r	r
S_p^2	Regression on	x_1 , $\cdots$, x_r , x_{r+1} , $\cdots$, x_p	p
$S_p^2 - S_r^2$	Difference		$p - r$
$S_T^2 - S_p^2$	Error		$N - p - 1$
S_T^2	Total		$N - 1$

TABLE 7

Analysis of Variance for $X = X_0$

		Sums of Squares	d.f.
$S_?^2$	Regression on X_0		1
S_p^2	Regression on x_1 , $\cdots$, x_p		p
$S_p^2 - S_1^2$	Difference		$p - 1$
$S_T^2 - S_p^2$	Error		$N - p - 1$
S_T^2	Total		$N - 1$

calculated formally from the regression of w on the sets x_1, $\cdots$, x_r and x_1, x_2, $\cdots$, x_p, with the analysis of variance given in Table 6.

Similarly, if we wish to test for the significance of a theoretical discriminant function, X_0, with preassigned coefficients, as compared with X_p, we have again the conventional test calculated from the formal analysis of the regression of w on x_1, x_2, $\cdots$, x_p, as given in Table 7.

As shown by Fisher [21] the relevant F-Test for this hypothesis is computable as

$$F_{p-1,n-p+1} = \frac{n - p + 1}{p - 1} \frac{R'^2}{1 - R'^2}$$

where $R'^2 = R^2(1 - r^2)$, r is the correlation between X and X_0 for fixed w, and R is the multiple correlation for w on $x_1, \cdots, x_p$, or, what is the same thing, the correlation of w and X.

The example of Smith [36] is an example in which the relationships of x's to w have to be estimated from analysis of variance and covariance of data in which the w's are not really known, being related to genotypes. The regression of x's on w is estimated by a generalization of the components-of-variance method, from variance-covariance analyses in which the usual null hypotheses are significantly contradicted. The net effect is that the usual significance tests now fail to hold, although the algebraic calculations are formally equivalent to those given above, once the population relations of x's to w are established. When work of this kind is based on small samples, there is some difficulty in estimating the reliability of the results.

3. Multi-dimensional discriminant functions. Instead of trying to discriminate between two populations or estimate a single parameter w, our problem may be to discriminate among several populations, not necessarily linearly related, or to estimate many independent parameters $w_1, w_2, \cdots, w_s$. Just as a single parameter w is sufficient to distinguish between means of measurements for two different populations, s parameters are sufficient to distinguish between means of $s + 1$ different populations, and exactly s parameters will be required, if no linear relation obtains among the $s + 1$ populations. For example, with three populations, any measurement mean may be given the three possible values α, $\alpha + \beta$, $\alpha + \gamma$, corresponding to $w_1 = w_2 = 0$ for population 1, $w_1 = 1$, $w_2 = 0$ for population 2, and $w_1 = 0$, $w_2 = 1$ for population 3. Geometrically we have to consider a set of parameter values as a point in an s-dimensional space.

The one-dimensional discriminant function admits two very different generalizations in higher dimensions. The practical solution to a particular problem for which s is moderately large may involve a mixture of both generalizations.

Let us generalize our statistical model before discussing the discrimination problem. To avoid complication of algebraic notation, let us for the moment assume $s = 2$. We will now postulate a set of hypothetical measurements $\xi_1, \xi_2, \cdots, \xi_p$, with

$$\xi_1 = \alpha_1 + \beta_1 u + \gamma_1 v + \epsilon_1$$

$$\xi_2 = \alpha_2 + \beta_2 u + \gamma_2 v + \epsilon_2$$

$$\xi_3 = \alpha_3 + \epsilon_3$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\xi_p = \alpha_p + \epsilon_p,$$

where the ϵ_p are independent normal deviates with unit variance, u and v are fixed variates or parameters corresponding to the different populations, and α_1, α_2, $\cdots$, α_p, β_1, β_2, γ_1, and γ_2 are constants. Evidently ξ_3, $\cdots$, ξ_p can yield no information about u and v; ξ_1 and ξ_2 together contain all the information there is to get about u and v. As before, assume that our data will be in the form of linear combinations $x_m = \Sigma l_{mn}\xi_n$, with unknown coefficients l_{mn}. The variance-covariance matrix within populations, or for fixed u, v, is still given by $\sigma_{mn} = \Sigma l_{mk}l_{nk}$. The mean values of the x's for fixed u, are given by

$$E(x_m) = \Sigma l_{mn}\alpha_n + (l_{m2}\beta_1 + l_{m2}\beta_2)u + (l_{m1}\gamma_1 + l_{m2}\gamma_2)v$$

$$= A_m + b_m u + c_m v.$$

This model is again justifiable on the basis of Hotelling's work.

The first question to ask is whether we can now form two linear combinations of the x's and get rid of ξ_3, $\cdots$, ξ_p in both, thus providing a two dimensional description of an individual on the basis of x_1, x_2, $\cdots$, x_p. The answer here is in the affirmative, as a result of a direct generalization of the method discussed earlier. If we calculate $X_1 = \Sigma \sigma^{mn} b_m x_n$ and $X_2 = \Sigma \sigma^{mn} c_m x_n$, we are fortunate enough to get

$$X_1 = \beta_1\xi_1 + \beta_2\xi_2$$

$$X_2 = \gamma_1\xi_1 + \gamma_2\xi_2$$

with no disturbing elements from ξ_3, $\cdots$, ξ_p. Assuming for now that X_1 and X_2 are not merely proportional, i.e. $\beta_1\gamma_2 - \beta_2\gamma_1 \neq 0$, what do we do with X_1 and X_2?

For fixed u, v, we have

$$E(X_1) = \Sigma \sigma^{mn} b_m a_n + u\Sigma \sigma^{mn} b_m b_n + v\Sigma \sigma^{mn} b_m c_n$$

$$= A_1 + B_1 u + C_1 v$$
$$E(X_2) = \Sigma \sigma^{mn} C_m a_n + u\Sigma \sigma^{mn} c_m b_n + v\Sigma \sigma^{mn} c_m c_n$$

$$= A_2 + B_2 u + C_2 v$$

and variances and covariance

$$\tau_{11} = \Sigma \sigma^{mn} b_m b_n = B_1$$

$$T_{12} = \Sigma \sigma^{mn} b_m c_n = C_1 = B_2$$

$$\tau_{22} = \Sigma \sigma^{mn} c_m c_n = C_2.$$

We may for example, estimate u and v by solving the equations

$$B_1 u + c_1 v = X_1 - A_1$$

$$B_2 u + C_2 v = X_2 - A_2,$$

or we may set up regions in the X_1, X_2 plane for which certain decisions are made. For example, when classifying an individual into one of three populations, we might delineate regions, as in Fig. 2.

Then the particular individual would be classified as coming from population I, II, or III, according to which region X_1, X_2 falls in. The individual points shown in the figure represent the expected values of X_1, X_2 for each of the three populations. No exhaustive investigation has been made for this situation, but some fairly obvious methods are available for constructing such regions.

With respect to significance tests when the σ_{mn}, a_m, b_m, c_m are estimated from samples, the whole gamut of multivariate analysis has to be run. Tests analogous to (but more complicated than) F tests exist for testing the significance

Classification Regions in X_1, X_2 Plane
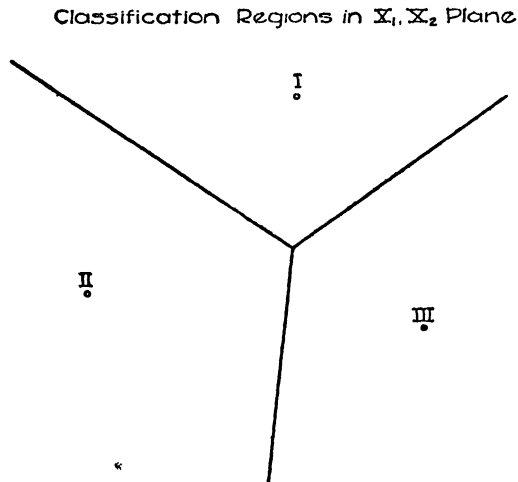
Fig. 2

of the discrimination, the significance of a subset of the x's, and the significance of a theoretical pair $X_{1,0}$, $X_{2,0}$ (Wilks [41], [42], [43]).

For some purposes a two-dimensional discrimant function X_1, X_2 may be unsatisfactory. For example, we might suspect that $\beta_1\gamma_2 = \beta_2\gamma_1$ (or that the relationship is nearly satisfied). Under these circumstances X_1 is (nearly) proportional to X_2, and we would like to compute the best one-dimensional discriminant function, even though we have started with two linear parameters u and v. Even if $\beta_1\gamma_2 \neq \beta_2\gamma_1$ we might still ask for the best one-dimensional discriminant function, in order to rank our populations on the "best" linear scale. If we define Y as that linear combination of x_1, x_2, $\cdots$, x_p which has the largest multiple correlation with u and v, we have generalized the simple one-dimensional discriminant function in a second direction.

Before proceeding, it is useful to recognize that Y, as defined above, must be a

function of X_1, X_2, since X_1 and X_2 together contain all the information about u and v that can be obtained from the x's.

Now suppose we consider an arbitrary linear combination $Y = \lambda_1 X_1 + \lambda_2 X_2$. Y correlates best with

$$\lambda_1(\tau_{11}u + \tau_{12}v) + \lambda_2(\tau_{12}u + \tau_{22}v) = (\lambda_1\tau_{11} + \lambda_2\tau_{12})u + (\lambda_2\tau_{12} + \lambda_2\tau_{22})v.$$

We now have to choose λ_1 and λ_2 to maximize this correlation. This correlation will be maximized if we maximize the ratio of the variance of

$$(\lambda_1\tau_{11} + \lambda_2\tau_{12})u + (\lambda_1\tau_{12} + \lambda_2\tau_{22})v$$

(over the distribution of u and v values) to the variance of Y for fixed u and v. Call the first quantity S_1, the second S_2. Then $S_2 = \lambda_1^2\tau_{11} + 2\lambda_1\lambda_2\tau_{12} + \lambda_2^2\tau_{22}$ and S_1 is of the form $\lambda_1^2\mu_{11} + 2\lambda_1\lambda_2\mu_{12} + \lambda_2^2\mu_{22}$ where

$$\mu_{11} = \tau_{11}^2\sigma_{uu} + 2\tau_{11}\tau_{12}\sigma_{uv} + \tau_{12}^2\sigma_{vv}$$

$$\mu_{12} = \tau_{11}\tau_{12}\sigma_{uu} + (\tau_{12}^2 + \tau_{11}\tau_{22})\sigma_{uv} + \tau_{12}\tau_{22}\sigma_{vv}$$

$$\mu_{22} = \tau_{12}^2\sigma_{uu} + 2\tau_{12}\tau_{22}\sigma_{uv} + \tau_{22}^2\sigma_{vv}.$$

Maximizing S_1/S_2 leads to the equations:

$$\lambda_1\tau_{11} + \lambda_2\tau_{12} = \frac{S_1}{S_2}(\lambda_1\mu_{11} + \lambda_2\mu_{12})$$

$$\lambda_1\tau_{12} + \lambda_2\tau_{22} = \frac{S_1}{S_2}(\lambda_1\mu_{12} + \lambda_2\mu_{22})$$

i.e.

$$\lambda_1(\tau_{11} - \theta\mu_{11}) + \lambda_2(\tau_{12} - \theta\mu_{12}) = 0$$

$$\lambda_1(\tau_{12} - \theta\mu_{12}) + \lambda_2(\tau_{22} - \theta\mu_{22}) = 0, \qquad \text{with } \theta = S_1/S_2.$$

It is thus seen that θ must satisfy the quadratic equation

$$(\tau_{11} - \theta\mu_{11})(\tau_{22} - \theta\mu_{22}) - (\tau_{12} - \theta\mu_{12})^2 = 0,$$

in order for solutions λ_1, λ_2 to exist. In general there will be two solutions, of which the greater corresponds to that linear combination $\lambda_1 X_1 + \lambda_2 X_2$ which has greatest multiple correlation with u and v, whereas the smaller corresponds to that linear combination which has least multiple correlation with u and v. θ itself corresponds to $R^2/(1 - R^2)$ for the regression of $\lambda_1 X_1 + \lambda_2 X_2$ on u, v.

In the general case with s degrees of freedom corresponding to w_1, w_2, $\cdots$, w_s, there is an s-dimensional discriminant function $(X_1, X_2, \cdots, X_s)$, and a set of s linear combinations for which $R^2/(1 - R^2)$ is stationary with respect to

$$\lambda_1, \cdots, \lambda_s.$$

The s roots (corresponding to an equation of degree s) arranged in decreasing order, permit construction of the best one-dimensional, two-dimensional, $\cdots$, $(s - 1)$-dimensional discriminant functions.

Discussion of the relevant significance tests for these reduced discriminant functions is beyond the scope of this paper. Reference may be made to the work of Hotelling and Fisher.

REFERENCES

[1] M. M. BARNARD, "The secular variations of skull characters in four series of Egyptian skulls", *Ann. Eugen.*, Vol. 6 (1935), pp. 352–371.

[2] M. S. BARTLETT, "The standard errors of discriminant function coefficients", *Jour. Roy. Stat. Soc.*, Suppl. 6 (1939), pp 169–173.

[3] M. S. BARTLETT, "Statistical significance of cannonical correlations", *Biometrika*, Vol. 32 (1942), pp. 29–37.

[4] W. D. BATEN, "The discriminant function applied to spore measurements", *Mich. Acad. of Sci., Arts, and Letters*, Vol. 29 (1943), pp. 3–7.

[5] W. D. BATEN AND C. C. DEWITT, "Use of the discriminant function in the comparison of proximate coal analyses", *Indust. and Eng. Chem., Anal. Ed.*, Vol. 16 (1944), pp. 32–34.

[6] W. D. BATEN AND H. M. HATCHER, "Distinguishing method differences by use of discriminant functions", *Jour. of Exp. Ed* , March, 1944.

[7] W. D. BATEN, "The use of discriminant functions in comparing judges' scores concerning potatoes", *Jour. Amer. Stat. Assoc.*, Vol. 40 (1945), pp. 223–227.

[8] R. C. BOSE, "On the exact distribution of D^2 statistic", *Sankhya*, Vol. 2 (1936), pp 143–154.

[9] R. C. BOSE AND S. N. ROY, "The exact distribution of the Studentized D^2 statistic", *Sankhya*, Vol. 4 (1938) pp. 19–38.

[10] G. W. BRIER, R. G. SCHOOT, AND V. L. SIMMONS, "The discriminant function applied to quality rating in sheep", *Proc. Amer. Soc. An. Prod.*, Vol. 1 (1940), pp. 153–160.

[11] W. G. COCHRAN, "The comparison of different scales of measurement for experimental results", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 205–216.

[12] G. M. COX AND W. P. MARTIN, "Use of a discriminant function for differentiating soils with different Azotobacter populations", *Iowa State Col. Jour. of Sci.*, Vol. 11 (1937), pp. 323–331.

[13] B. B. DAY AND M. M. SANDOMIRE, "Use of the discriminant function for more than two groups", *Jour. Amer. Stat. Assoc.*, Vol. 37 (1942), pp. 461–472.

[14] W. E. DEMING, "On the chi test and curve fitting", *Jour. Amer. Stat. Assoc.*, Vol. 29 (1934), pp. 372–382.

[15] D. DURAND, "Risk elements in consumer installment financing", *Nat. Bur. Econ. Res., Inc.*, Studies in Consumer Installment Financing, No. 8, 1941.

[16] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient", *Proc. Roy. Soc. A.*, Vol. 121 (1928), pp. 654–673.

[17] R. A. FISHER, *"Statistical Methods for Research Workers"*, Oliver and Boyd, Section 29.

[18] R. A. FISHER, "The use of multiple measurements in taxonomic problems", *Ann. Eugen.*, Vol. 7 (1936), pp. 179–188.

[19] R. A. FISHER, "Statistical utilization of multiple measurements", *Ann. Eugen.*, Vol. 8 (1938), pp. 376–386.

[20] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations", *Ann. Eugen.*, Vol. 9 (1939), pp. 238–249.

[21] R. A. FISHER, "The precision of discriminant functions", *Ann. Eugen.*, Vol. 10 (1940), pp. 422–429.

[22] H. E. GARRETT, "The discriminant function and its use in psychology", *Psychometrika*, Vol. 8 (1943), pp. 65–79.

[23] M. A. GIRSHICK, "On the sampling theory of the roots of determinantal equations", *Annals of Math. Stat.*, Vol. 10 (1939), pp. 203–224.

[24] T. HAALVELMO, "Probability applications in econometrics", *Econometrica*, Suppl. Vol. 12 (1944).

[25] H. HOTELLING, "Generalization of Student's ratio", *Annals of Math. Stat.* ,Vol. 2 (1931), pp. 360–378.

[26] H. HOTELLING, "The most predictable onterion", *Jour. Educ. Psychol.*, Vol. 26 (1935), pp. 139–142.

[27] H. HOTELLING, "Relations between two sets of variates", *Biometrika*, Vol. 28 (1936), pp. 321–377.

[28] P. L. HSU, "On the distribution of roots of certain determinantal equations", *Ann. Eugen.*, Vol. 9 (1939), pp. 250–258.

[29] W. G. MADOW, "Contributions to the theory of multivariate statistical analyses", *Trans. Am. Math. Soc.*, Vol. 44 (1938), pp. 454–495.

[30] P. C. MAHALANOBIS, "Analysis of race mixture in Bengal", *Jour. Asiat. Soc. Beng.*, Vol. 23 (1927), pp. 301–333.

[31] P. C. MAHALANOBIS, "On tests and measures of group divergence", *Jour. Asiat. Soc. Beng.*, Vol. 26 (1930), pp. 541–588.

[32] P. C. MAHALANOBIS, "On the generalized distance in statistics", *Proc. Mat. Inst. Sci. Ind.*, Vol. 12 (1936), pp. 49–55.

[33] E. A. MARTIN, "A study of an Egyptian series of mandibles, with special reference to mathematical methods of sexing", *Biometrika*, Vol. 28 (1936), pp. 149–178.

[34] K. PEARSON, "On the coefficient of racial likeness", *Biometrika*, Vol. 18 (1926) ,pp. 105–117.

[35] C. R. RAO, "Tests with discriminant functions in multivariate analysis", *Sankhya*, Vol. 7 (1946), pp. 407–414.

[36] H. F. SMITH, "A discriminant function for plant selection", *Ann. Eugen.*, Vol. 7 (1936), pp. 240–250.

[37] W. L. STEVENS, "The standardization of rubber flexing tests", *India Rubber World*, August 1, 1940.

[38] A. WALD, "On a statisical problem arising in the classification of an individual into one of two groups", *Annals of Math. Stat.*, Vol. 14 (1944), pp. 145–162.

[39] N. WALLACE AND R. M. W. TRAVERS, "A psychometric sociologia study of a group of specialty salesmen", *Ann. Eugen.*, Vol. 8 (1938), pp. 266–302.

[40] F. V. WAUGH, "Regression between sets of variables", *Econometrica*, Vol. 10 (1942), pp. 290–310.

[41] S. S. WILKS, "Certain generalizations in the analysis of variance", *Biometrika*, Vol. 24 (1932), pp. 471–494.

[42] S. S. WILKS, "On the independence of *k* sets of normally distributed statistical variables", *Econometrica*, Vol. 3 (1935), pp. 309–326.

[43] S. S. WILKS, *Mathematical Statistics*. Princeton Univ. Press, 1943.

NON-PARAMETRIC ESTIMATION II. STATISTICALLY EQUIVALENT BLOCKS AND TOLERANCE REGIONS—THE CONTINUOUS CASE

By John W. Tukey

Princeton University

1. Summary. Wald [2, 1943] extended the usefulness of tolerance limits to the simplest multi-dimensional cases. His principle is here used to provide many new ways of using a sample of n to divide the range of the population into $n + 1$ blocks of known behavior. The exact tolerance distribution for the proportions of the population covered by these blocks is extended from the case of a continuous probability density function to the case of a continuous cumulative distribution function. Such an extension is needed in dealing completely with multivariate cases *even* where the underlying distribution is as smooth as a multivariate normal distribution.

The devices used in Paper I [1] to extend the usefulness of tolerance limits to the case of a discontinuous underlying distribution will be applied in the next paper of this series, with some extension, to extend the usefulness of these general tolerance regions to the case of a discontinuous distribution. Some of these results specialize into new results for the univariate case, although they do not seem to have any immediate practical application.

The author wishes to acknowledge the stimulation given to his work on this problem by Henry Scheffé, whose modesty has kept this paper from the joint authorship of papers I [1, Scheffé and Tukey 1945] and IV (not yet written).

2. Introduction. Wald's great contribution to the theory of tolerance limits was his method of successive elimination. As originally presented for a bivariate situation it ran roughly as follows: Let (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) be a sample of n from an arbitrary bivariate population. The type of tolerance region to be used is determined by four preassigned integers, k_1, k_2, k_3, and k_4. The procedure is as follows: Order the n observations according to their x values. Select the k_1 highest, and let the x coordinate of the lowest of these k_1 be x_u. Select the k_2 lowest, and let the x coordinate of the highest of these k_2 be x_l. Discard these $k_1 + k_2$ selected observations, and order the remaining $n - k_1 - k_2$ observations according to their y values. Select the k_3 highest of these remaining observations, and let the y coordinate of the lowest of these k_3 be y_u. Select the k_4 lowest of these remaining observations, and let the y coordinate of the highest of these k_4 be y_l. The tolerance region, consisting of all points (x, y), with $x_l < x < x_u$ and $y_l < y < y_u$ depends on the sample, and, hence, so does the fraction of the population falling in (= covered by) this region. Wald showed that the distribution of this fraction covered was independent of the underlying bivariate distribution, so long as this latter distribution had a continuous probability density function. He showed that the

distribution was the same as that arising in the one-dimensional case when a tolerance region was set with the aid of $k_1 + k_2 + k_3 + k_4$ observations. (Numerical approximation to these distributions will be discussed in Paper IV of this series.

The important device in this process, and the one which makes the conclusion possible, is the discarding of the $k_1 + k_2$ observations after they have played their part by determining x_l and x_u.

We shall shortly be able to describe this procedure of Wald's as a special case of a more general procedure, but we shall first go back to the simplest one dimensional case to explain some of our notions and terminology.

Consider the uniform distribution from 0 to 1, draw a sample of n, and let the sample values, ordered according to size be t_1, t_2, $\cdots$, t_n. These n values divide the interval from 0 to 1 into the following $n + 1$ parts $(0, t_1)$, (t_1, t_2), $\cdots$, $(t_{n-1}t_n)$, $(t_n, 1)$ which we shall call *blocks*. Since the joint distribution of the t_i is well known, that of the lengths of these $n - 1$ blocks is easily found. This distribution of lengths would be unimportant, if it were not at the same time the distribution of the fractions of the population covered by the blocks. As is shown later, this distribution of fractions covered, or, more simply, of *coverages*, has the following properties:

(i) the fractions covered add up to 1.

(ii) the distribution is completely symmetrical.

Property (ii) makes intuitive the result of Wilks [3, 1941] that the distributions of the coverage of regions obtained

(a) by removing the $k_1 + k_2$ left-most blocks,

(b) by removing the k_1 left-most and the k_2 right-most blocks

are identical. The specific distribution obtained satisfies

(iii) if the coverages are taken as barycentric coordinates on an n-simplex, the distribution over the simplex is uniform,

(iv) the sum of the coverages of any k *pre*selected blocks of the $n + 1$ has the well-known distribution

$$Pr \{\text{sum of } k \text{ coverages} < t\} = I_t (n - k + 1, k)$$

where $I_\beta (n, m)$ is the incomplete Beta function.

We shall call a set of blocks, derived from a sample, whose coverages behave in this general way a set of *statistically equivalent blocks*. Normally this will be abbreviated to se-blocks. (A precise definition is given in section 4.)

We shall concentrate much of our attention on all the blocks and their symmetrical character, rather than on the tolerance region formed by deleting k of them, since our results will then be applicable to many other problems.

Now we can generalize Wald's original procedure. Let W_1, W_2, $\cdots$, W_n be a sample of n—we shall not need to consider its distribution—and let φ_1, φ_2, $\cdots$, φ_n be n numerically valued functions of W, possibly alike, possibly distinct, such that $\varphi_1(W), \varphi_2(W), \cdots, \varphi_n(W)$ have a joint distribution. Proceed as follows:

Order the W_i according to the numbers $\varphi_1(W_i)$, select the W_i for which $\varphi_1(W_i)$ is largest and denote it by $W_{i(1)}$. The first block contains all W such that

(2.1a)
$$\varphi_1(W) > \varphi_1(W_{i(1)}).$$

Discarding $W_{i(1)}$, order the remaining W_i according to the values of $\varphi_2(W_i)$, and select as $W_{i(2)}$ the one giving the largest value. The second block contains all W such that

(2.1b)
$$\varphi_1(W) < \varphi_1(W_{i(1)}),$$
$$\varphi_2(W) > \varphi_2(W_{i(2)}).$$

Continue this process. The mth block, for $m \leq n$ will be defined by

(2.1m)
$$\varphi_j(W) < \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, m-1,$$
$$\varphi_m(W) \gtrless \varphi_m(W_{i(m)}),$$

and the $(n + 1)$st block by

(2.1n)
$$\varphi_j(W) < \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, n.$$

(A graphical example of this construction is given shortly.) This set of $n + 1$ blocks will be statistically equivalent whenever the cumulative distribution of each φ_i function is continuous.

To specialize this to the case described above, let W be a pair (x, y) of numbers and let

  (i)   the first $k_1 \varphi$'s be the x-coordinate of W,
  (ii) the next $k_2 \varphi$'s be *minus* the x-coordinate of W,
  (iii) the next $k_3 \varphi$'s be the y-coordinate of W,
  (iv) the next $k_4 \varphi$'s be *minus* the y-coordinate of W,
  (v) the remaining φ's be arbitrary.

Then the first k_1 blocks will contain all W for which

$$x = \varphi_j(W) > \varphi_j(W_{i(j)}), \qquad j = 1, 2, \cdots, k_1$$

that is, for which

$$x > x_u = \varphi_{k_1}(W_{i(k_1)}).$$

Similarly, the next $k_2 + k_3 + k_4$ blocks will contain all W with

$$x < x_l,$$
$$y > y_u, \qquad x_l \leq x \leq x_u,$$
$$y < y_l, \qquad x_l \leq x \leq x_u,$$

respectively, and the removal of these $k_1 + k_2 + k_3 + k_4$ blocks leaves Wald's tolerance region (plus the boundaries where $x = x_u$, $x = x_l$, $y = y_u$, $y = y_l$).

There would be no point in this more general wording, if it did not include

new cases of some interest.   We give now, in graphic terms, an example of such a case.

We deal with a sample of n bivariate observations, which we think of as plotted on a *map* so that we can use geographical language.   The number n is rather large, and we wish to construct a tolerance region by deleting 12 blocks.   We proceed as follows:

Find the most northerly point, draw an East-West line through it, and shade the area North of the line.   Find the most easterly point in the unshaded area, 'r w a North-South line through it, and shade the unshaded area East of the
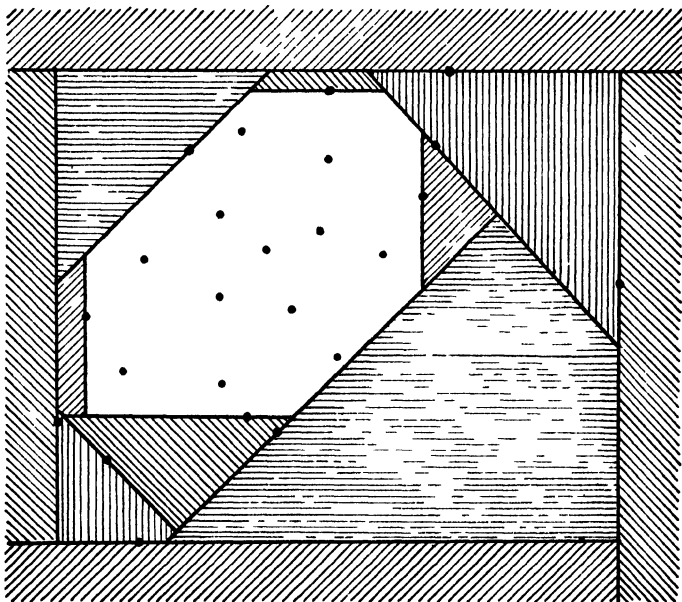
Fig. 1

ine.   Find the most southerly point, (always working in the unshaded area), draw an East-West line through it and shade the area South of the line.   Find the most westerly point, draw a North-South line through it, and shade the area West of the line.   Find the most northeasterly point, draw a NW-SE line through it and shade the area northeast of the line.   Find the most southeasterly point, draw a NE-SW line through it, and shade the area southeast of the line.   Repeat this 6 times more, choosing in succession the most southwesterly, northwesterly, northerly, easterly, southerly, and westerly points.   The remaining points will now lie in an unshaded area surrounded by a polygon, which will have 8 (or perhaps fewer) sides.   The inside of this polygon is the desired tolerance region.

Figure 1 shows the final result, starting from $n = 25$. The practicing statistician is invited to try an example of his own with n at least 100.

Other newly accessible cases can easily be invented by the reader, after he considers this example carefully.

The use of a single W and n functions φ_i has two virtues; it simplifies notation and frees the intuition, as compared with the use of n chance quantities $Z_i = \varphi_i(W)$.

If the bivariate situation above were regarded as a 12-variate situation, where the variates were, in order, $(y, x, -y, -x, x + y, x - y, -x - y, -x + y, y, x, -y, -x)$ then the original Wald procedure with $k_1 = k_3 = \cdots = k_{23} = 1$; $k_2 = k_4 = \cdots = k_{24} = 0$ would apply to construct the same region. Yet even if x and y had a bivariate normal distribution, Wald's proof would not apply without extension. For the 12-dimensional distribution is highly singular (it is concentrated on a 2-dimensional plane in 12-dimensional space) and there is no hope of a density function. An extension of Wald's result to the case where the 12-dimensional joint cumulative distribution function is continuous —as is the case in this example when x and y have a continuous joint cumulative —is clearly needed.

When we come to deal with the case of where the cumulative needs not be continuous we shall meet a further difficulty, namely "ties". But if, as in the present case, the cumulative is continuous, it is easy to see that the probability that $\varphi_i(W_j) = \varphi_i(W_k)$ for any i, j, k is zero.

3. Terminology and notation. A quantity which has a probability distribution we call a *chance quantity* (it has frequently been called a *random variable*). The term chance quantity does not imply that its values are single real numbers, they may be single real numbers (when we also speak of a real chance quantity), sets of n real numbers, or more general objects. The cumulative distribution function, or *cumulative*, of a single real chance quantity, X, is defined by

$$F(t) = Pr\{X < t\},$$

except perhaps at the discontinuities of F. We have used here the notation $Pr\{k(X)\}$ to indicate the probability that $k(X)$ holds, and we have followed our policy of using capital letters for chance quantities and the corresponding lower case letters for their values.

The set of values of W, or, as we shall say, the W-set, for which, for example $\varphi(W) \leq 3$, will be denoted by

$$\{W \mid \varphi(W) \leq 3\}.$$

We shall wish to compute probabilities associated with one or more functions of a chance quantity; usually we will emphasize that these functions shall be measurable with respect to the probability measure underlying the distribution of W by asserting that they have a joint cumulative, which is defined by

$$F(t_1, t_2, \cdots, t_k) = Pr\{\varphi_k(W) < t_k\},$$

(except possibly at discontinuities of F) and which does not exist unless the φ_i are measurable with respect to the unknown underlying distribution of W. In cases where we neglect to remind the reader, it is still assumed that the functions are measurable.

The coverage of a W-set, which may itself be a chance quantity, is defined by

$$\text{Coverage of } S = Pr\ \{W \ \epsilon \ S\}.$$

When S is a chance quantity, its coverage is also a chance quantity. The barycentric simplex (of dimension n) is the set of points in $n + 1$-dimensional Euclidean space $(t_1, t_2, \cdots, t_{n+1})$ with $t_1 + t_2 + \cdots + t_{n+1} = 1$ and $0 \le t_i \le 1$. The name comes from the representation of the point $(t_1, t_2, \cdots, t_{n+1})$ as the center of gravity (in mechanical terms) or mean (in statistical terms) of the distribution where a fraction t_i is concentrated at the ith vertex. (In order, the vertices are $(1, 0, 0, \cdots, 0)$, $(0, 1, 0, \cdots, 0)$, etc.) The uniform distribution on this simplex has an (n-dimensional) density

$$n!\,dt_1 dt_2 \cdots dt_n, \qquad (0 \le t_1, t_2, \cdots, t_n, 1 - t_1 - t_2 \cdots - t_n \le 1),$$

and the cumulative

$$T(x_1, x_2, \cdots, x_{n+1}) = n! \int\int \cdots \int dt_1 dt_2 \cdots dt_n$$

where the integration is over the range where $0 \le t_i \le x_i$ *and at the same time* $t_1 + t_2 + \cdots + t_{n-1} \le 1$.

4. The blocks determined by n values of W. We deal now with a population of W's (a probability measure μ on the space $T \equiv \{w\}$), a family of functions $\varphi_1, \varphi_2, \cdots, \varphi_m$ of W with a joint cumulative (measurable with respect to μ) and a set of values $w_1, w_2, \cdots, w_n$, $(w_i \ \epsilon \ T)$.

(4.1) DEFINITION *The set $w_1, w_2, \cdots, w_n$ and the functions $\varphi_1, \varphi_2, \cdots, \varphi_m$ define blocks as follows:*

$$(4.2) \qquad\qquad S_1 = \{w \mid \varphi_1(w) > a_1\}$$

where $a_1 = \max_i \varphi_1(w_i) = \varphi_1(w_{i(1)})$, which defines $i(1)$.

$$(4.3) \qquad\qquad S_2 = \{w \mid \varphi_1(w) < a_1, \varphi_2(w) > a_2\},$$

where $a_2 = \max_{i \neq i(1)} \varphi_2(w_i) = \varphi_2(w_{i(2)})$, $i(2) \neq i(1)$, which defines $i(2)$. And in general, for $1 < k \le \min(m, n)$,

$$(4.4) \qquad S_k = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_{k-1}(w) < a_{k-1}, \varphi_k(w) > a_k\},$$

where $a_k = \max'_i \varphi_k(w_i) = \varphi_k(w_{i(k)})$, *the maximum being taken over all* i *except* $i(1), i(2), \cdots, i(k-1)$; *and* $i(k)$ *being chosen distinct from all* $i(j), j < k$.

If $m \geq n$, *then*

(4.5) $$S_{n+1} = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_n(w) < a_n\}.$$

If $m \leq n$, *then*

(4.6) $$S_{m \mid n+1} = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_m(w) < a_m\}.$$

The result of this definition is to use $w_1, \cdots, w_n$ and $\varphi_1, \cdots, \varphi_m$ to define $n + 1$ blocks (one more than there are w's) in case there are enough functions, and, in case there are not enough functions, to define one small block, S_i, for each function plus one large remainder $S_{m \mid n+1}$. We notice

(4.2) REMARK. *The blocks of* (4.1) *are well defined unless* $\varphi_1(w_j) = \varphi_i(w_k)$ *for some* i, j, k.

5. Statement of results for the statistician.

The central results can be stated as follows:

(5.1) THEOREM $A_{m \mid n+1}$. *If* $W_1, W_2, \cdots, W_n$ *are a sample of* n *from a distribution, if* $\varphi_1, \varphi_2, \cdots, \varphi_m$, $(m \leq n)$, *are* m *functions such that*

$$\varphi_1(W), \varphi_2(W), \cdots, \varphi_m(W)$$

have a joint distribution which has a continuous cumulative, and if the blocks $S_1, S_2, \cdots, S_m$ *and* $S_{m \mid n+1}$ *are defined as in* (4.1), *then*

(i) *the blocks are disjoint chance sets, uniquely defined with probability one,*
(ii) *the distribution of the coverages*

$$c_i = Pr\{w \text{ in } S_i\}, \qquad i = 1, 2, \cdots, m$$

and

$$c_{m \mid n+1} = Pr\{w \text{ in } S_{m \mid n+1}\}$$

is the same as that of $t_1, t_2, \cdots, t_m$ *and* $t_{m+1} + t_{m+2} + \cdots + t_{m+1}$ *where* t_i *are uniformly distributed on the barycentric simplex with* $n + 1$ *vertices.*

Conditions (5.1i) and (5.1ii) are the precise definition of *a partial family of statistically equivalent blocks of type* $n + 1$ *and an associated* $(m \mid n + 1)$ *tolerance region.*

(5.2) THEOREM B_{n+1}. *If* $W_1, W_2, \cdots, W_n$ *are a sample of* n *from a distribution, and if* $\varphi_1, \varphi_2, \cdots, \varphi_m$, $(m \geq n)$, *are* m *functions such that*

$$\varphi_1(W), \varphi_2(W), \cdots, \varphi_m(W)$$

have a joint distribution which has a continuous cumulative, and if the blocks $S_1, S_2, \cdots, S_{n+1}$ *are defined as in* (4.1), *then*

(i) *the blocks are disjoint chance sets, defined with probability one.*

(ii) *the distribution of the coverages*

$$c_i = Pr \{w \text{ in } S_i\}, \qquad i = 1, 2, \cdots, n + 1$$

is the same as that of t_1, t_2, $\cdots$, t_{n+1}, where the t_i are uniformly distributed on the barycentric simplex with $n + 1$ vertices.

Conditions (5.2i) and (5.2ii) are the precise definition of *a complete family of statistically equivalent blocks.* In Paper III we shall have to widen these notions a little, and this form will then be qualified by the phrase "in the narrow sense".

6. Statement of results for the measure theorist. The construction of (4.1) maps the product $T^n \times U^n$ into E^{n+1} where T is the set of w's (and hence T^n is the set of ordered n-tuples of w's), U is the space of all real-valued functions defined over T, measurable with respect to a fixed probability measure μ, and possessing a continuous cumulative, (i.e. $\mu(\{w \mid \varphi(w) = c\}) = 0$ for all real c), and hence U^n is the space of ordered n-tuples of such functions, and E^{n+1} is Euclidean n-dimensional space. More precisely, the mapping is into the barycentric simplex with $n + 1$ vertices, a subset of E^{n+1}, and is well defined except for a set in T^n of measure zero with respect to μ^n, the power measure of μ. In these terms, we may restate theorem B as follows:

(6.1) Theorem B_{n+1}. *Hold the n functions φ_1, φ_2, $\cdots$, φ_n and the probability measure fixed, then T^n is mapped into B_n and the power measure μ^n is carried by that mapping into a measure on B_n. This measure is always $n!$ times Lebesgue measure.*

7. Wald's principle. The essential principle behind Wald's process of discarding observations is sufficiently fundamental to warrant a name of its own. It can be stated, quite generally, in the two following forms:

(7.1) Wald's Principle. (discrete form.) *Let W be a chance quantity, and consider samples of n. Fix disjoint w-sets A_1, A_2, $\cdots$, A_m, B. Consider those samples of n for which exactly one value falls in each A_i and the remaining $n-m$ fall in B. The distribution of the $n-m$ falling in B is that of a random sample of $n-m$ from the distribution of W restricted to B.* (i.e. $\mu_B(X) = [\mu(B)]^{-1}\mu(BX)$.)

(7.2) Wald's Principle. (conditional form.) *Let W be a chance quantity, and φ a function such that each value of $\varphi(W)$ has probability zero. Consider samples of n. Then the conditional distribution of the w_i, given that*

$$\max_i \varphi(w_i) = a,$$

is that of one w_{i_0} with $\varphi(w_{i_0}) = a$ and a sample of $n-1$ other w_i from the distribution of W restricted to $B = \{w \mid \varphi(w) < a\}$.

(7.3) Central Lemma. *Let W be a chance quantity and let φ_1, $\cdots$, φ_n be functions with a joint cumulative such that $\varphi_i(w) = a$ has probability zero for each i and a (i.e. the joint cumulative is continuous). Then the conditional distribu-*

tion of the remaining $n-k$ w's, after k blocks have been chosen according to (4.1) *is that of a sample from the distribution of W restricted to*

$$B = \{w \mid \varphi_1(w) < a_1, \cdots, \varphi_k(w) < a_k\},$$

where $k = 1, 2, \cdots, n$.

The proofs of these statements are elementary and direct. To establish (7.1) we have only to show that given two sets in B^{n-k}, their probabilities on the assumption that one w_i is in each A_i are in the ratio of their probabilities for an unrestricted sample of $n-k$. But the probability of finding the $n-k$ w_i in a set R, contained in B^{n-k}, and one w_i in each A_i, is exactly

$$\frac{n!}{(n-k)!} \mu(A_1)\mu(A_2) \cdots \mu(A_k)$$

times the probability that $n - k$ w_i, known to be in B^{n-k}, will fall in R. This establishes (7.1).

In order to prove (7.2) we must show that the probability of a set R of n-tuples $w_1, w_2, \cdots, w_n$ is the same whether calculated directly or calculated by the proposed conditional distribution. To this end, it is natural to decompose R as follows:

$$R = R(1) + R(2) + \cdots + R(n) + Z,$$

where $R(i)$ contains those $(w_1, \cdots, w_n)$ in R for which $\varphi(w_i) > \varphi(w_j)$ for all $j \neq i$, and Z contains the remaining $(w_1, \cdots, w_n)$; which must involve at least one tie $\varphi(w_j) = \varphi(w_k)$, $j \neq k$. Since Z has probability zero, it will suffice to establish the equality of the two calculations for sets of the form $R(i)$, and because of symmetry we may restrict ourselves to sets of the form $R(1)$.

Given an integer N, we decompose the range of $\varphi(w)$ into Nn segments of equal probability, which we may do because the cumulative of φ is continuous. There are then Nn values b_k, $(b_o = -\infty, b_{Nn} = +\infty)$ such that

$$Pr \{b_{k-1} < \varphi(w) < b_k\} = 1/Nn.$$

We now decompose our set R (which is of the form $R(1)$ as follows:

$$R = R_2 + \cdots + R_{Nn} + Y,$$

where R_k contains those n-tuples

$$(w_1, \cdots, w_n) \text{ for which } b_{k-1} < \varphi(w_1) < b_k$$

and $\varphi(w_i) < b_{k-1}$ for all $i > 1$. The remaining set Y contains n-tuples where the two largest $\varphi(w_i)$, $(i = 1 \text{ and } i = i_o)$, belong to the same interval. The probability of this is less than

$$\frac{n(n-1)}{2}\left(\frac{1}{nN}\right)^2 \le \frac{1}{2N^2}$$

as calculated from the known distribution. Calculating from the conditional distribution, we find immediately a bound of

$$\sum_k \left\{ \left(\frac{k}{Nn}\right)^n - \left(\frac{k-1}{Nn}\right)^n \right\} \frac{n-1}{k} = \frac{n-1}{(Nn)^n} \sum \frac{k^n - (k-1)^n}{k}$$

$$= \frac{n-1}{(Nn)^n} \left((nN)^{n-1} + \sum_{k<nN} \frac{k^{n-1}}{k-1} \right)$$

$$\leq \frac{n-1}{(Nn)^n} \left(A_n (nN)^{n-1} \right)$$

$$= \frac{A_n}{N},$$

where A_n is a constant depending only on n. Thus, as N increases, the probability of the successive sets Y tend to zero—calculated either way. To show the equivalence of the two calculations it is now sufficient to show that they agree for the sets R_k. But this is a case of (7.1) and the lemma is proved.

Now (7.3) follows by induction, applying (7.2) at each step.

8. Proof of theorems. We notice that Theorem B_n is equivalent to Theorem $A_{m|n+1}$, since, according to (4.1) $S_{n|n+1} = S_{n+1}$.

We have only to prove theorem $A_{m|n+1}$, which we do by induction on m. For $m = 1$, it is exactly Wilks' [3, 1941] original one-dimensional theorem, and is known. Let us assume it for $m = k$ and demonstrate it for $m = k + 1$, for by induction this will complete the proof.

We must deal with the blocks $S_1, S_2, \cdots, S_k, S_{k+1}$ and $S_{k+1|n+1}$, (notation as in (4.1) and (5.1)). We need the obvious

(8.1) **LEMMA.** *Since the cumulative of φ_{k+1} is continuous, the union of S_{k+1} and $S_{k+1|m+1}$ differs from $S_{k|m+1}$ by a set of zero probability.*

Hence

$$c_{k|n+1} \equiv c_{k+1} + c_{k+1|n+1}.$$

Since we know from the induction hypothesis that $c_1, c_2, \cdots, c_k$ and $c_{k|n+1}$ have the correct joint distribution, we have only to show that c_{k+1} and c_1, $c_2, \cdots, c_k$ have the correct joint distribution. Fix $c_1, c_2, \cdots, c_k$. Then $a_1, a_2, \cdots, a_k$ must be fixed, and so (7.3) applies to the $n-k$ w_i's not discarded after $a_1, a_2, \cdots, a_k$ have been fixed. The conditional distribution of c_{k+1} must be that of a fixed number $(1 - c_1 - c_2 - \cdots - c_k)$, which is the probability attached to $S_{k|n+1}$, times the coverage of one block based on a sample of $n-k$, since the remaining $n-k$ w's behave like a sample.

Consider the very particular case where w is uniformly distributed between zero and one and $\varphi_i(w) \equiv w$, all that we have said in the last paragraph applies —the conditional distribution of c_{k+1} given $c_1, c_2, \cdots, c_k$ is the same in the two cases—hence the joint distribution of $c_1, c_2, \cdots, c_k, c_{k+1}$ is the same in both cases—but in this very particular case the joint distribution is known to be that required by theorem $A_{k+1|n+1}$.

REFERENCES

[1] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192. (Also cited as Paper I.)

[2] A. Wald, "An extension of Wilks' method for setting tolerance limits", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[3] S. S. Wilks, "Determination of sample sizes for setting tolerance limits", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 91–96.

SOME BASIC THEOREMS FOR DEVELOPING TESTS OF FIT FOR THE CASE OF THE NON-PARAMETRIC PROBABILITY · DISTRIBUTION FUNCTION, I

By Bradford F. Kimball

State Department of Public Service, New York, N. Y.

1. Summary. In developing tests of fit based upon a sample $O_n(x_i)$ in the case that the cumulative distribution function $F(X)$ of the universe of X's is not necessarily a function of a finite number of specific parameters—sometimes known as the non-parametric case—it has been pointed out by several writers that the "probability integral transformation" is a useful device (cf. [1]–[4]).

The author finds that a modification of this approach is more effective. This modification is to use a transformation of ordered sample values x_i from a random sample $O_n(x_i)$ based on successive *differences* of the cdf values $F(x_i)$.

A theorem is proved giving a simple formula for the expected values of the products of powers of these differences, where all differences from 1 to $n + 1$ are involved in a symmetrical manner.

The moment generating function of the test function defined as the sum of m squares of these successive differences is developed and the application of such a test function is briefly discussed.

2. Introduction. Let the sample values x_i be ordered so that

$$(2.1) \qquad\qquad x_i \leq x_{i+1}, \qquad\qquad (i = 1, 2, \cdots, n - 1).$$

Let F_r denote the value of the cdf $F(X)$ associated with the rth ordered sample value x_r. Thus

$$(2.2) \qquad\qquad F_r = F(x_r).$$

Consider the following transformation of the ordered sample values x_i based upon the (hypothetically) known cumulative distribution function $F(X)$ which will be taken as a continuous function of X over its admissible range:

$$(2.3) \qquad \begin{aligned} u_1 &= F_1, \\ u_r &= F_r - F_{r-1}, \\ u_{n+1} &= 1 - F_n. \end{aligned} \qquad\qquad (r = 2, 3, \cdots, n)$$

The restrictions on F_i are that

$$(2.4) \qquad\qquad F_i \leq F_{i+1}, \text{ and } 0 \leq F_i \leq 1.$$

The above transformation (2.3) translates these conditions into the *symmetrical* conditions

$$(2.5) \qquad 0 \leq u_i, \text{ and } u_1 + u_2 + \cdots + u_n + u_{n+1} = 1.$$

A one-to-one correspondence between u_i and F_i exists if one of the u_i be omitted,—say u_β. With u_β omitted, the Jacobian of the transformation from F_i to u_i

540

has value unity. The probability density of the sample $O_n(x_i)$, with x_i ordered, is given by

$$(2.6) \qquad P[O_n(x_i)] \, dO_n = n! \, dF_1 \, dF_2 \cdots dF_n \, .$$

Hence with u_β omitted,

$$(2.7) \qquad P[O_n(x_i)] \, dO_n = n! \, du_1 \, du_2 \cdots du_{\beta-1} \, du_{\beta+1} \cdots du_{n+1} \, .$$

The sample space of the u_i with u_β omitted, is that portion of the $n + 1$ Euclidean space of all the u_i variables, bounded by the coordinate hyperplanes, which is on the projection of the hyperplane (2.5) upon the hyperplane $u_\beta = 0$. This is a region in the n-space of the u_i with u_β omitted, bounded by the coordinate hyperplanes and the hyperplane

$$(2.8) \qquad u_1 + u_2 + \cdots + u_{\beta-1} + u_{\beta+1} + \cdots + u_n + u_{n+1} = 1.$$

Thus the formal integral of the pdf of the u_i over sample space is

$$(2.9) \qquad n! \int n \int du_1 \cdots du_{\beta-1} \, du_{\beta+1} \cdots du_{n+1} = 1$$

with $0 \leq u_i$, and u_i bounded above by the hyperplane (2.8).

It is now clear that both the pdf and the sample space of the u_i (with u_β omitted) are symmetrical in the u_i. This fact leads to *complete* symmetry of the joint distribution function of any set of u_i, over $i = 1$ to $n + 1$ including u_β, relative to the u_i selected. Other interesting results are forthcoming.

3. Basic mathematical theorem. Using the techniques associated with the Beta function, the expectation of the products of powers u_i is found to be

$$(3.1) \qquad \begin{aligned} &E[u_r^p \cdot u_s^q \cdot u_t^w \cdots] \\ &= \Gamma(n+1) \Gamma(p+1) \Gamma(q+1) \Gamma(w+1) \cdots / \Gamma(n+p+q+w+\cdots+1) \end{aligned}$$

where r, s, t, etc., are any set of different indices (for the present other than β) from the integers 1 to $n + 1$, and p, q, w, etc., are any real numbers greater than minus one. The relation (3.1) can further be generalized to the case where u_β may be included. This will be proved for the case $n = 2$, with p, q and w taken as integers. The generalization can be concluded from inspection. Thus with

$$u_3 = 1 - u_1 - u_2,$$

$$\begin{aligned} E[u_1^p \cdot u_2^q \cdot u_3^w] &= 2! \int_0^1 u_2^q \, du_2 \int_0^{1-u_2} u_1^p (1 - u_1 - u_2)^w \, du_1 \\ &= 2! \int_0^1 u_2^q (1 - u_2)^{p+w+1} \int_0^1 s^p (1 - s)^w \, ds \\ &= \frac{2! \, p! \, w!}{(p + w + 1)!} \int_0^1 u_2^q (1 - u_2)^{p+w+1} \, du_2 = \frac{2! \, p! \, q! \, w!}{(p + q + w + 2)!} \, . \end{aligned}$$

Hence the theorem:

THEOREM. *Given a random sample of n values of X from a universe with cdf $F(X)$ which is continuous over the range of X. With the sample values x_i ordered so that $x_i \leq x_{i+1}$ define a set of $n + 1$ variables u_i as the successive differences of $F(x_i)$ by the relations (2.3). The expected value of the product of real powers greater than minus one of any or all of the u_i, $(i = 1, 2, \cdots, n + 1)$, is given by the relation (3.1) above (not subject to the omission of u_β).*

There are many interesting consequences of this theorem. Perhaps the most striking is the following:

COROLLARY 1. *Let a range $\alpha(m, k)$ for positive integer m be defined by*

$$(3.2) \qquad \alpha(m, k) = F(x_{k+m}) - F(x_k)$$

with $\qquad k = 0, 1, 2, \cdots, n, \text{ and } m \leq n + 1 - k$

under the convention

$$F(x_0) = 0, \qquad F(x_{n+1}) = 1.$$

The probability distribution of $\alpha(m, k)$ is independent of k and hence is the same as that of $F(x_m)$.

Another interesting consequence (not new) is the following:

COROLLARY 2. *The correlation of u_i and u_k, $i \neq k$, is the same for all pairs (i, k) over the range of indices from 1 to $n + 1$, and has the value $-1/n$.*

Introducing the notation

$$(3.3) \qquad [n + r]_r = (n + r)(n + r - 1) \cdots (n + 1),$$

the corollary follows from the relationships

$$E(u_i) = 1/(n + 1), \quad E(u_i^2) = 2/[n + 2]_2, \quad E(u_i u_k) = 1/[n + 2]_2.$$

The fact that the correlation between any two frequency differences u_i and u_k is negative leads to the following more general relationship:

COROLLARY 3. *For any set of different indices i, j, k, etc., and for any positive numbers p, q, r, etc., the expectation of the product of the powers p, q, r, $\cdots$ of u_i, u_j, $u_k \cdots$ is less than the product of the expectations of the powers taken separately:*

$$(3.4) \qquad E[u_i^p \cdot u_j^q \cdot u_k^r \cdots] < E(u_i^p) \cdot E(u_j^q) \cdot E(u_k^r) \cdots.$$

This follows from generalization of the relation

$$\frac{\Gamma(n + 1)\Gamma(p + 1)\Gamma(q + 1)\Gamma(r + 1)}{\Gamma(n + p + q + r + 1)}$$

$$< \frac{[\Gamma(n + 1)]^3 \Gamma(p + 1)\Gamma(q + 1)\Gamma(r + 1)}{\Gamma(n + p + 1)\Gamma(n + q + 1)\Gamma(n + r + 1)}.$$

The above theorem suggests the possibility of test functions for fitted distributions, relative to a universe with a cdf which, since it is merely conditioned by a sufficient hypothesis for the theorem, may be of the non-parametric type.

A test function of the form

(3.5) $$Y = \sum_m u_i^p, \qquad p \text{ real and positive}$$

might first come to mind. If $p = 1$, compensatory effects of deviations reduce the efficiency of the test function. One is thus led first to consider the test function (3.5) for the case $p = 2$.

4. The moments of the probability distribution of $y_m = \Sigma u_i^2$. We are first concerned with the problem of the determination of the moments of the function

(4.1) $$y_m = \sum_m u_i^2$$

where i ranges over any particular fixed set of m integers which for simplicity is usually taken as the first m.

One first recalls the fact that the result is independent of *which* m indices have been selected; and that the expected value of any combination of powers is independent of which specific subscripts of u_i are involved.

Since the u_i are correlated, principles of combinatory analysis are involved in determining the moments of y_m. One possible way of obtaining the moments is as follows:

Let v_r denote the rth moment of y_m about $y_m = 0$. Thus

(4.2) $$E[(y_m)^r] = v_r \doteq E[(\sum_m u_i^2)^r].$$

Now in the expansion of $(\sum_m u_i^2)^r$, the sum of the power indices of each term is $2r$. Thus referring back to (3.1) and (3.3) it will be noted that the expected value of each such term will have the common factor

$$1/[n + 2r]_{2r}.$$

Consider a general term of the expansion of $(\sum_m u_i^2)^r$

$$C_{r_1 r_2 \cdots r_k} \cdot u_{i_1}^{2r_1} u_{i_2}^{2r_2} \cdots u_{i_k}^{2r_k}, \qquad \text{with} \quad r_1 + r_2 + \cdots + r_k = r.$$

Clearly

$$E(u_{i_1}^{2r_1} u_{i_2}^{2r_2} \cdots u_{i_k}^{2r_k}) = 2r_1 ! \, 2r_2 ! \cdots 2r_k !/[n + 2r]_{2r}.$$

and the coefficient $C_{r_1 r_2 \cdots r_k}$ is the multinomial coefficient

$$C_{r_1 r_2 \cdots r_k} = \frac{r!}{r_1! r_2! \cdots r_k!}.$$

Now in the expansion of $(\sum_m u_i^2)^r$ group the terms which have the same set of k values of r_i, irrespective of which indices of u_i are involved. The number of such terms (since each involves k different indices) is $\binom{m}{k}$. If $r_1, r_2, \cdots, r_k$,

are all different each combination could be taken in $k!$ different ways. Thus with r's all different and fixed, the sum of all coefficients of terms with same combination of $2r_i$ powers (irrespective of variation of indices of the u_i) is

$$\binom{m}{k} k! \frac{r!}{r_1! r_2! \cdots r_k!}.$$

This would then constitute the total multiplier for

$$2r_1 ! \, 2r_2 ! \cdots 2r_k !/[n + 2r]_{2r}$$

for a given set of k r's which are all different.

If some of r's are repeated, let $k_1, k_2, \cdots, k_s$ denote the number of repetitions of each different r_i ($k_i \geqq 1$, and $k_1 + k_2 + \cdots + k_s = k$). Then each combination of the k r's corresponding to a set of k products could be taken in

$$k!/(k_1! \, k_2! \cdots k_s!)$$

different ways. Hence the lemma:

LEMMA 1. *Consider all admissible sets of k different subscripts of u_i and a fixed set of values of $r = r_1, r_2, \cdots, r_k$ where*

$$r_1 + r_2 + \cdots + r_k = r$$

such that s of these r's are different, and the number of repetitions in the set of r's is given by $k_1 k_2 \cdots k_s$ ($k_i \geqq 1$, and $k_1 + k_2 + \cdots + k_s = k$). The composite coefficient of the terms in v_r involving the factor

$$2r_1! \, 2r_2! \cdots 2r_k!/[n + 2r]_{2r}$$

is given by

(4.3)
$$\binom{m}{k} \frac{k!}{k_1! k_2! \cdots k_s!} \cdot \frac{r!}{r_1! r_2! \cdots r_k!}.$$

Examples of computation of v_r by means of the above lemma. The first order moment is given by

(4.4)
$$v_1 = E(\textstyle\sum u_i^2) = m \, 2!/[n + 2]_2.$$

The second order moment is given by

$$v_2 = E[(\textstyle\sum u_i^2)^2] = C_1 E(u_i^4) + C_2 E(u_i^2 u_j^2),$$

and determining the values of C_i from Lemma 1,

$$v_2 = \left[m4! + \binom{m}{2}\left(\frac{2!}{2!}\right)\frac{2!}{1!1!} 2!2! \right] \Big/ [n + 4]_4$$

or

(4.5)
$$v_2 = \left[m4! + 8\binom{m}{2} \right] \Big/ [n + 4]_4 = \left[m + \binom{m}{2}\frac{1}{3} \right] \Big/ \binom{n + 4}{4}.$$

Again for the third order moment,

$$v_3 = E[(\sum_m u_i^2)^3] = C_1 E(u_i^6) + C_2 E(u_i^2 u_j^4) + C_3 E(u_i^2 u_j^2 u_k^2),$$

and using Lemma 1,

$$= \left[m6! + \binom{m}{2} \frac{2!}{1!1!} \frac{3!}{1!2!} 2!4! + \binom{m}{3} \frac{3!}{3!} \frac{3!}{1!1!1!} 2!2!2! \right] \Big/ [n+6]_6$$

$$= \left[m6! + \binom{m}{2} 2!3!4! + \binom{m}{3} 2!2!2!3! \right] \Big/ [n+6]_6$$

or

(4.6)
$$v_3 = \left[m + \binom{m}{2} \frac{2}{5} + \binom{m}{3} \frac{1}{15} \right] \Big/ \binom{n+6}{6}.$$

Similarly writing the fourth moment in the form

$$v_4 = C_1 E(u_i^8) + C_2 E(u_i^6 u_j^2) + C_3 E(u_i^4 u_j^4) + C_4 E(u_i^2 u_j^2 u_k^4) + C_5 E(u_i^2 u_j^2 u_k^2 u_i^2)$$

and using Lemma 1 it reduces to

(4.7)
$$v_4 = \left[m + \binom{m}{2} \frac{2}{7} + \binom{m}{2} \frac{3}{35} + \binom{m}{3} \frac{3}{35} + \binom{m}{4} \frac{1}{105} \right] \Big/ \binom{n+8}{8}.$$

. Higher order moments of the probability distribution function may be computed as desired.

An alternate method of computing the moments of the distribution of this test function is the following:

Consider a function $g_0(x)$ such that

(4.8)
$$\frac{d^r g_0(0)}{dx^r} = (2r)!, \qquad g_0(0) = 1.$$

Thus

(4.9)
$$E[u^{2r}] = [d^r g_0(0)/dx^r]/[n+2r]_{2r}.$$

From the principles of combinatory analysis of linear operators, it follows that[1]

(4.10)
$$E[(\sum_m u_i^2)^r] = \frac{d^r [g_0(x)]^m}{dx^r} \Big|_{x=0} /[n+2r]_{2r}.$$

Although this is an enlightening analytical form, actual computations seem to be simpler with the use of Lemma 1.

---

[1] One way of seeing this is to first think of the u_i as statistically independent. The numerators of the resulting terms would be the same as in (4.10). When the u_i are taken as dependent, by virtue of (3.1) the numerators will remain the same while all denominators will reduce to $[n + 2r]_{2r}$.

Moment generating function. The moment generating function of the probability distribution of y_m can be written as

$$(4.11) \quad E(e^{ty}) = G_0(t, m) = 1 + \sum_{r=1}^{\infty} [d^r(g_0(x))^m/dx^r \mid _{x=0}]/[n + 2r]_{2r} \, t^r/r!$$

with

$$g_0(x) = 1 + 2 \,!\, x + 4 \,!\, x^2/2! + 6 \,!\, x^3/3! + \cdots + (2r) \,!\, x^r/r! + \cdots$$

$$[n + 2r]_{2r} = (n + 2r)(n + 2r - 1) \cdots (n + 1).$$

Although $g_0(x)$ exists only as a formal power series, $G_0(t, m)$ is defined by (4.11) as a power series with positive coefficients, converging for all t.

5. Some comments on test function, $p = 2$. At the present time the study of the test function for $p = 2$ has not gone far enough to justify publication of results. One difficulty is that although its asymptotic distribution function appears to be normal, the convergence towards normalcy may be extremely slow in some cases.

Furthermore there are indications that the case $m = n + 1$ will give the most definitive results not only because the complete range of data is used, but also because errors of Type II would in general have a less erratic effect.

For the case $m = n + 1$ the mean, variance and third and fourth reduced moments (i.e. moments about the mean divided by corresponding power of σ) are:

Case $m = n + 1$.

$$E(y_{n+1}) = 2/(n + 2), \qquad \sigma^2 = 4n/[(n + 2)^2(n + 3)(n + 4)],$$

$$\alpha_3 = \mu_3/\sigma^3 = \frac{10n - 4}{(n + 5)(n + 6)} \sqrt{\frac{(n + 3)(n + 4)}{n}}$$

(5.1)

$$\alpha_4 = \left[\frac{n^3 + 101n^2 + 14n - 8}{(n + 5)(n + 6)(n + 7)(n + 8)} \right]\left[\frac{3(n + 3)(n + 4)}{n} \right]$$

$$\alpha_4 - 3 = \frac{6(41n^4 + 241n^3 + 118n^2 - 784n - 48)}{n(n + 5)(n + 6)(n + 7)(n + 8)}.$$

If data is not grouped the test may be applied as follows: Given a function $Q(X)$ which has been fitted to the cdf $F(X)$. From a random sample of size n with x_i ordered as in (2.1) compute the successive differences of $Q(x_i)$ to obtain the variables u_i^*. Then consider the sum of the squares

$$U^* = \sum_{n+1} u_i^{*2}.$$

If $Q(X)$ is a true representation of $F(X)$ the variation of U^* will follow that of y_{n+1}. Thus the expected value of U^*, its variance etc. will be *independent* of the fitted function $Q(X)$, which represents certain advantages over the χ^2 test.

The effect of Type II errors can be roughly analyzed as follows: In considering the effect of such errors the testing procedure must be criticized from the point of view that

$$Q(X) \neq F(X).$$

For $m = n + 1$ it still is true that

$$\Sigma u_i^* = 1$$

which tends to act as a control upon U^*. For example set

$$u_i^* = u_i + \chi_i .$$

Then from the above relation it follows that

(5.2) $$\Sigma \chi_i = 0.$$

Write U^* as

(5.3)
$$U^* = \Sigma u_i^2 + \Sigma \chi_i^2 + 2\Sigma u_i \chi_i$$
$$= \Sigma u_i^2 + \Sigma \chi_i^2 + (2\Sigma \chi_i)/(n + 1) + 2\Sigma \chi_i \delta(u_i)$$

where $\delta(u_i)$ denotes the variation of the true frequency differences from their expected value $1/(n + 1)$.

The variation $\delta(u_i)$ will be to a considerable degree independent of χ_i. Thus the term $\Sigma \chi_i^2$ will in general tend to be larger than the last term on the right. The third term on the right will be zero by virtue of (5.2), and hence U^* will tend to be larger than y_{n+1}. A similar effect upon the sampling variance of U^* can be noted. Hence an interval of rejection

$$U^* \geqq A, \qquad P[y_{n+1} \leqq A] = \alpha = \text{confidence level,}$$

is pointed to.

On the other hand if $m < n + 1$ the condition (5.2) no longer holds, the term $(2\Sigma \chi_i)/(n + 1)$ of (5.3) will not be zero and in many cases would dominate the other two error terms. Thus it is easily conceivable that one may have in the case $m < n + 1$

$$U_m^* < y_m$$

even when the discrepancies χ_i are large. Hence in the case $m < n + 1$ choice of confidence interval will require considerable care (see [1]).

Although the distribution of y_{n+1} for small n is decidedly non-normal, if the test function is replaced by

(5.4) $$r_{n+1} = (\Sigma[u_i - 1/(n + 1)]^2)^{\frac{1}{2}}$$

it will be found that the probability density function takes on the normal character quite rapidly with increasing n. Indeed the author has found that a computed approximation to the probability density function of r_{n+1} with $n = 4$ is decidedly normal in character.

REFERENCES

[1] J. NEYMAN, "Smooth test for goodness of fit," *Skand. Aktuar. Tidskn.* (1937) p. 149.

[2] E. S. PEARSON, "The probability integral transformation for testing goodness of fit and combining independent tests of significance," *Biometrika*, Vol. 30 (1938), pp. 134–148.

[3] E. J. GUMBEL, "Simple tests for given hypothesis," *Biometrika*, Vol. 32 (1942), pp. 317–333.

[4] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation, I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.

AN ESSENTIALLY COMPLETE CLASS OF ADMISSIBLE DECISION FUNCTIONS

By Abraham Wald

Columbia University

Summary. With any statistical decision procedure (function) there will be associated a risk function $r(\theta)$ where $r(\theta)$ denotes the risk due to possible wrong decisions when θ is the true parameter point. If an a priori probability distribution of θ is given, a decision procedure which minimizes the expected value of $r(\theta)$ is called the Bayes solution of the problem. The main result in this note may be stated as follows: Consider the class C of decision procedures consisting of all Bayes solutions corresponding to all possible a priori distributions of θ. Under some weak conditions, for any decision procedure T not in C there exists a decision procedure T^* in C such that $r^*(\theta) \leqq r(\theta)$ identically in θ. Here $r(\theta)$ is the risk function associated with T, and $r^*(\theta)$ is the risk function associated with T^*. Applications of this result to the problem of testing a hypothesis are made.

1. Introduction. In some previous publications [1], [2] the author has considered the following general problem of statistical inference: Let $X = (X_1, \cdots, X_n)$ be a set of chance variables. Suppose that the only information we have concerning the joint distribution function F of these chance variables is that F is an element of a given class Ω of distribution functions. Suppose, furthermore, that a class D of possible decisions d is given one of which is to be made on the basis of an observation $x = (x_1, \cdots, x_n)$ on the chance vector X. The problem is then to construct a function $d(x)$, called statistical decision function, which associates with each sample point x an element $d(x)$ of D so that the decision $d(x)$ is made when the sample point x is observed. A statistical decision function $d(x)$ is defined over all possible points x of the sample space and for each sample point x the value of the function is an element of D. Each element d of D will usually be interpreted as a decision to accept the hypothesis that the unknown distribution F of X belongs to a certain subclass ω of Ω. Different elements d of D correspond to different subclasses ω of Ω.

The problem of testing the hypothesis H that the unknown distribution function F belongs to a given subclass ω of Ω, is contained as a special case in the above general problem. The space D will then contain only two elements, d_1 and d_2, where d_1 denotes the decision of accepting H and d_2 denotes the decision of rejecting H.

As in [1] and [2], we shall assume also here that Ω is a k-parameter family of distribution functions. Then each element of Ω may be represented by a point $\theta = (\theta_1, \cdots, \theta_k)$, called parameter point, in the k-dimensional Cartesian space. The class Ω is then represented by a subset of the k-dimensional Cartesian space,

called parameter space. We shall, therefore, refer to Ω as the parameter space and to its elements as parameter points.

The merits of any particular decision function $d(x)$ will usually depend on the relative importance of the various possible errors caused by not selecting the proper element d of D. The relative importance of such errors has been described in [1] and [2] by a weight function $W(\theta, d)$ defined over the product of Ω and D. For any pair (θ, d) the value of $W(\theta, d)$ is non-negative and expresses the loss caused by taking the decision d when θ is the true parameter point. For any given decision function $d(x)$ the expected value of the loss is given by

$$(1.1) \qquad r(\theta) = \int_M W[\theta, d(x)] \, dF(x)$$

where M denotes the sample space and $F(x)$ is the joint cumulative distribution of $X = (X_1, \cdots, X_n)$ corresponding to the parameter point θ.

The function $r(\theta)$ is defined over the parameter space Ω and is called the risk function. The shape of the risk function $r(\theta)$ will, in general, be affected by the decision function $d(x)$ used. To put this dependence in evidence, we shall use the symbol $r[\theta \mid d(x)]$ to denote the risk function $r(\theta)$ associated with the decision function $d(x)$.

A decision function $d(x)$ is said to be uniformly better than the decision function $d^*(x)$ if

$$(1.2) \qquad r[\theta \mid d(x)] \leq r[\theta \mid d^*(x)]$$

for all θ and if there exists at least one point θ for which the inequality sign holds in (1.2). A decision function $d(x)$ is said to be admissible if no other uniformly better decision function exists.

A class C of admissible decision functions will be said to be essentially complete if for any decision function $d(x)$ not in C there exists a decision function $d^*(x)$ in C such that

$$r[\theta \mid d^*(x)] \leq r[\theta \mid d(x)]$$

for all θ.

In section 2 we shall formulate certain assumptions which will then be used in section 3 to derive an essentially complete class of admissible decision functions. In section 4 applications are made to the problem of testing a hypothesis.

In a recent paper Lehmann [3] obtained an essentially complete class of admissible tests for each hypothesis H of a certain restricted class of simple hypotheses. The restrictions imposed on Ω in Lehmann's paper are essentially those formulated by Neyman [4], [5] to insure the existence of the type A_1 (uniformly most powerful unbiassed) test. Our definition of an essentially complete class of admissible decision functions agrees with that given by Lehmann when the problem is to test a hypothesis and the weight function $W(\theta, d)$ can take only the values 0 and 1.

2. Assumptions. Throughout this paper we shall make the following assumptions:

Assumption 1: The parameter space Ω is a bounded and closed subset of a finite dimensional, say k-dimensional, Cartesian space.

We shall introduce the following convergence definition in the space D: a sequence $\{d_m\}$, $(m = 1, 2, \cdots, \text{ad inf.})$, of elements of D is said to converge to the element d of D if

$$\lim_{m \to \infty} W(\theta, d_m) = W(\theta, d)$$

uniformly in θ.

Assumption 2: The space D is compact and, for any d, $W(\theta, d)$ is a continuous function of θ.

Assumption 3: For any point θ of Ω the joint distribution function of $X = (X_1, \cdots, X_n)$ admits a density function $p(x, \theta)$ for all points x of the n-dimensional Cartesian space M (sample space). The density function $p(x, \theta)$ is assumed to be continuous in x and θ jointly.

In what follows we shall mean by a distribution function $f(\theta)$ of θ a cumulative distribution function for which $\int_\Omega df(\theta) = 1$ and for which $\int_\Omega W(\theta, d) df(\theta)$ is not zero identically in d.

Assumption 4: For any point x of M, except perhaps for a set of measure zero, and for any cumulative distribution function $f(\theta)$ there exists one and only one element of D for which the expression

$$(2.1) \qquad \int_\Omega W(\theta, d) p(x, \theta) \, df(\theta)$$

takes its minimum value with respect to d.

Assumptions 1 and 3 in this paper are exactly the same as Assumptions 1 and 3 in [2]. The formulation of Assumptions 2 and 4 is somewhat different from that given in [2]. This is mainly due to the fact that in [2] the space D has the same elements as Ω, while here this is not necessarily so. It can be verified without difficulty that this slight modification of the assumptions does not affect in any way the validity of the results obtained in [2]. Thus, we shall be able to make use of any theorems proved in [2] for the purposes of the present paper.

3. Derivation of an essentially complete class of admissible decision functions. For any distribution function $f(\theta)$ defined over Ω and for any sample point x let $d(x, f)$ denote the element of D for which the expression (2.1) takes its minimum value. It follows easily from the definition of $r(\theta)$ and $d(x, f)$ that

$$(3.1) \qquad \int_\Omega r[\theta \mid d(x, f)] df(\theta) \leq \int_\Omega r[\theta \mid d^*(x)] \, df(\theta)$$

for any decision function $d^*(x)$. If we interpret $f(\theta)$ as an a priori probability distribution of θ, inequality (3.1) says that the expected value of $r(\theta)$ takes its minimum value for the decision function $d(x, f)$. We shall refer to $d(x, f)$ as the Bayes' solution of the problem corresponding to the a priori probability distribution $f(\theta)$.

We shall now prove the following theorem.

THEOREM 3.1. *The class C of all Bayes' solutions $d(x, f)$ corresponding to all possible a priori distributions $f(\theta)$ is an essentially complete class of admissible decision functions.*

PROOF. First we show that for any distribution $f(\theta)$ the decision function $d(x, f)$ is admissible. Let $d(x)$ be a decision function such that

$$r[\theta \mid d(x)] \leqq r[\theta \mid d(x, f)]$$

for all θ. Then

$$(3.2) \qquad \int_\Omega r[\theta \mid d(x)] \, df(\theta) \leqq \int_\Omega r[\theta \mid d(x, f)] \, df(\theta).$$

From the definition of $d(x, f)$ it follows that the equality sign must hold in (3.2), i.e.,

$$(3.3) \qquad \int_\Omega r[\theta \mid d(x)] \, df(\theta) = \int_\Omega r[\theta \mid d(x, f)] \, df(\theta).$$

From the second half of Theorem 4.2 in [2] it then follows that

$$r[\theta \mid d(x)] = r[\theta \mid d(x, f)]$$

for all θ. Hence $d(x, f)$ is an admissible decision function.

We shall now show that the class C of decision functions $d(x, f)$ corresponding to all possible a priori distributions $f(\theta)$ is essentially complete. Let $d_0(x)$ be any decision function not in the class C. The essential completeness of the class C is proved if we can show that there exists a distribution $f(\theta)$ such that

$$(3.4) \qquad\qquad r[\theta \mid d(x, f)] \leqq r[\theta \mid d_0(x)]$$

for all θ.

To prove (3.4) we shall consider the weight function

$$(3.5) \qquad W^*(\theta, d) = W(\theta, d) - r[\theta \mid d_0(x)] + \underset{\theta}{\text{Max}} \, r[\theta \mid d_0(x)]$$

The maximum of $r[\theta \mid d_0(x)]$ exists, since according to Theorem 4.1 in [2] $r[\theta \mid d_0(x)]$ is a continuous function of θ. Clearly, Assumptions 1–4 remain valid if we replace $W(\theta, d)$ by $W^*(\theta, d)$. Let $r^*[\theta \mid d(x)]$ denote the risk function associated with the decision function $d(x)$ if the weight function is given by $W^*(\theta, d)$. According to Theorem 5.2 in [2] there exists a decision function $d^*(x)$ such that

$$(3.6) \qquad\qquad \underset{\theta}{\text{Max}} \, r^*[\theta \mid d^*(x)] \leqq \underset{\theta}{\text{Max}} \, r^*[\theta \mid d(x)]$$

for any decision function $d(x)$. Since

$$\text{Max}_\theta r^*[\theta \mid d_0(x)] = \text{Max}_\theta r[\theta \mid d_0(x)]$$

it follows from (3.6) that

(3.7) $$\text{Max}_\theta r^*[\theta \mid d^*(x)] \leq \text{Max}_\theta r[\theta \mid d_0(x)].$$

Inequalities (3.5) and (3.7) imply

(3.8) $$r[\theta \mid d^*(x)] \leq r[\theta \mid d_0(x)]$$

for all θ.

For any distribution $f(\theta)$ we shall denote by $d^*(x, f)$ the Bayes solution of the problem corresponding to the a priori distribution $f(\theta)$ when the weight function is given by $W^*(\theta, d)$. Since $W^*(\theta, d) - W(\theta, d)$ depends only on θ but not on d, one can easily verify that $d^*(x, f) = d(x, f)$. It follows from Theorems 4.4 and 5.1 in [2] that there exists a distribution $f(\theta)$, the so-called least favorable distribution, such that (3.6) remains valid if we replace $d^*(x)$ by $d^*(x, f)$. Thus we can put

(3.9) $$d^*(x) = d^*(x, f) = d(x, f).$$

Hence, from (3.8) we obtain

$$r[\theta \mid d(x, f)] \leq r[\theta \mid d_0(x)]$$

for all θ. This completes the proof of Theorem 3.1.

4. Applications to the problem of testing a hypothesis. In this section we shall apply the results of the preceding section to the problem of testing the hypothesis H that the true parameter point is included in a given subset ω of Ω. We shall assume that ω is an open subset of Ω. The space D consists now only of two elements, d_1 and d_2, where d_1 denotes the decision of accepting H and d_2 denotes the decision of rejecting H.

We shall assume that the $W(\theta, d_1)$ is equal to zero for points θ in the interior or on the boundary of ω, and positive elsewhere. Similarly, $W(\theta, d_2)$ will be assumed to be positive for points θ inside ω and zero outside ω. For any a priori distribution $f(\theta)$ the Bayes solution is given by the following test: We reject the hypothesis H if (and only if)[1]

(4.1) $$\int_{\Omega-\omega} W(\theta, d_1)p(x, \theta) \, df(\theta) > \int_\omega W(\theta, d_2)p(x, \theta) \, df(\theta).$$

Thus, the class C of regions (4.1), corresponding to all possible distributions $f(\theta)$, is an essentially complete class of admissible critical regions.

For any critical region R we shall denote the probability that the sample x

---

[1] Whether the equality sign is included or not in (4.1) is of no consequence, since by Assumption 4 the measure of the set of points x for which the equality holds in (4.1) is zero.

will fall in R when θ is true by $P(\theta \mid R)$. It follows from Lemma 4.4 in [2] and Assumption 3 that $P(\theta \mid R)$ is a continuous function of θ for any region R. Since $W(\theta, d_1)$ is positive in the interior of $\Omega - \omega$, and $W(\theta, d_2)$ is positive in ω, the class C of regions defined in (4.1) will have the following properties:

(a) For any region R outside the class C there exists a region R^* in C such that

$$P(\theta \mid R^*) \leqq P(\theta \mid R) \text{ in } \omega$$

and

$$P(\theta \mid R^*) \geq P(\theta \mid R) \text{ in } \Omega - \omega.$$

(b) If R and R^* are members of C such that

$$P(\theta \mid R^*) \leqq P(\theta \mid R) \text{ in } \omega$$

and

$$P(\theta \mid R^*) \geq P(\theta \mid R) \text{ in } \Omega - \omega,$$

then

$$P(\theta \mid R^*) = P(\theta \mid R) \text{ for all } \theta.$$

For any distribution $g(\theta)$ consider the critical region consisting of all sample points x satisfying

$$(4.2) \qquad \int_{\Omega-\omega} p(x, \theta) \, dg(\theta) > \int_{\omega} p(x, \theta) \, dg(\theta).$$

Let C^* be the class of regions (4.2) corresponding to all possible distributions $g(\theta)$. One can easily verify that any region in C is also a member of C^*. Thus, the following theorem holds:

THEOREM 4.1   *Suppose that Assumptions 1 and 3 are fulfilled and ω is an open subset of Ω. Suppose, furthermore, that for any distribution $g(\theta)$ the set of sample points x satisfying the equation*

$$\int_{\Omega-\omega} p(x, \theta) \, dg(\theta) = \int_{\omega} p(x, \theta) \, dg(\theta)$$

has the measure zero. Then, for any region R outside the class C^* there will be a region R^* in C^* such that

$$P(\theta \mid R^*) \leq P(\theta \mid R) \text{ in } \omega$$

and

$$P(\theta \mid R^*) \geqq P(\theta \mid R) \text{ in } \Omega - \omega.$$

Addition at proof reading: After this paper was sent to the printer, the author obtained a generalization of Theorem 3.1 to sequential decision functions, as well as some other results.   They will appear in a forthcoming issue of *Econometrica.*

REFERENCES

[1] A. WALD, "Contributions to the theory of statistical estimation and testing hypotheses", *Annals of Math. Stat.*, Vol. 10 (1939).

[2] A. WALD, "Statistical decision functions which minimize the maximum risk", *Annals of Math.*, Vol. 46 (1945).

[3] E. L. LEHMANN, "On families of admissible tests", *Annals of Math. Stat.*, Vol. 18 (1947), pp. 97–104.

[4] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part II", *Statistical Research Memoirs*, Vol. II, London, 1938.

[5] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses, Part I", *Statistical Research Memoirs*, Vol. I, London, 1936.

DISCRIMINATING BETWEEN BINOMIAL DISTRIBUTIONS

By Paul G. Hoel

University of California at Los Angeles

1. Summary. Given a set of k random samples, x_1, x_2, $\cdots$, x_k, from a binomial distribution with parameters p and n, it is shown that the familiar binomial index of dispersion

$$z = \frac{\sum_1^k (x_i - \bar{x})^2}{\bar{x}\left(1 - \frac{\bar{x}}{n_0}\right)}$$

yields an approximate best critical region independent of p for testing the hypothesis $n = n_0$ against the alternative hypothesis $n > n_0$, provided $\bar{x}$ and $n_0 - \bar{x}$ are not small. Because of the nature of the test, its optimum properties also apply to testing whether the data came from a binomial population with $n = n_0$ or from a Poisson population.

2. Introduction. A problem of considerable interest in certain fields is that of deciding whether a set of observations should be treated as having come from either a binomial population or from a Poisson population. Although there was much discussion a few years ago concerning the best method for making such a decision [1], [2], [3], no solution of the problem was presented. In this paper a test that possesses certain optimum properties is derived for discriminating between two binomial populations. This test, however, is also capable of solving the problem of how to discriminate between a binomial and a Poisson population. The methods that are employed in the derivation of this test are similar to those of an earlier paper [4] in which the problem of discriminating between two Poisson populations was studied.

3. Similar regions. Let n denote the number of trials and p the probability of success in a single trial for a binomial distribution. Let x_1, x_2, $\cdots$, x_k represent the observed frequencies in k random samples from this binomial population. Now consider the two alternative hypotheses

$$H_0 : n = n_0 , p = p_0$$

and

$$H_1 : n = n_1 > n_0 , p = p_1.$$

The purpose of this paper is to construct a test for discriminating between the two values of n regardless of the values of p; however it is convenient to begin with these more restrictive hypotheses

For the purpose of finding a critical region for testing H_0 against H_1, the x_i will be treated as the coordinates of a point in k dimensions. The probability of obtaining the particular point $x_1, \cdots, x_k$ when H_0 is true will be denoted by $P_0[x_i]$. Since the probability of obtaining x successes in n trials is given by

$$\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

it follows that

(1) $$P_0[x_i] = \frac{(n_0!)^k}{\prod_1^k x_i!(n_0 - x_i)!} p_0^{\sum_1^k x_i} q_0^{\sum_1^k (n_0 - x_i)}$$

In searching for a critical region that will be independent of p_0, it is illuminating to study the methods that were designed by Neyman and Pearson [5] for continuous distributions. These methods suggest that one should look for critical regions on the surfaces $\sum_1^k x_i = $ constant. For this reason, instead of using (1) for constructing critical regions, it is desirable to study the conditional probability distribution of the points lying in the plane $\sum_1^k x_i = N$, where N is a positive integer not exceeding kn_0. The conditional probability of obtaining the point $x_1, \cdots, x_k$, when the point is restricted to lie in the plane $\sum_1^k x_i = N$, will be denoted by $P_0[x_i \mid N]$. Its value may be obtained by dividing the probability (1) by the probability that the point will lie in the plane $\sum_1^k x_i = N$. If this latter probability is denoted by $P_0[N]$, then

(2) $$P_0[x_i \mid N] = \frac{P_0[x_i]}{P_0[N]}.$$

Since the sum of k independent variables each possessing the same binomial distribution has a binomial distribution with n replaced by kn, it follows that N possesses a binomial distribution and that

(3) $$P_0[N] = \frac{(kn_0)!}{N!(kn_0 - N)!} p_0^N q_0^{kn_0 - N}.$$

If (1) and (3) are substituted in (2), it will reduce to

(4) $$P_0[x_i \mid N] = \frac{(n_0!)^k N!(kn_0 - N)!}{(kn_0)! \prod^k x_i!(n_0 - x_i)!}$$

This conditional probability distribution in the plane $\sum_1^{\tilde{}} x_i = N$ is independent of p_0 and therefore may serve as the basis for constructing a critical region that

is independent of p_0 for testing H_0 against H_1. It will therefore be possible to test the less restrictive hypothesis

$$H_0' : n = n_0$$

against

$$H_1' : n = n_1 > n_0 .$$

4. Best critical region. Although a best critical region does not exist for testing H_0' against H_1', it is helpful to proceed as though one did.

If a critical region of size α could be selected in each plane $\sum_1^k x_i = N$, ($N = 0, 1, \cdots, kn_0$), then the totality of such critical regions would constitute a critical region of size α that is independent of p_0 and which therefore could be used to test H_0' against H_1'. For, if $P_0[X \in \text{C.R.}]$ denotes the probability that the sample point, which will be denoted by X, will lie in the critical region, it follows that

$$P_0[X \in \text{C.R.}] = \sum_{N=0}^{kn_0} P_0[N]P_0[X \in \text{C.R.} \mid N]$$

(5)
$$= \sum_{N=0}^{kn_0} P_0[N]\alpha$$

$$= \alpha.$$

This last equality follows from the fact that the sample point must lie in one of the planes $\sum_1^k x_i = N$, ($N = 0, 1, \cdots, kn_0$).

Furthermore, this would be the only critical region of size α independent of p_0, because if a critical region of size α_N, ($N = 0, 1, \cdots, kn_0$), were selected in the plane $\sum_1^k x_i = N(N = 0, 1, \cdots, kn_0)$, it would be necessary that

$$\sum_{N=0}^{kn_0} P_0[N]\alpha_N = \alpha,$$

independent of the value of p_0. From (3) this is equivalent to requiring that

(6)
$$\sum_{N=0}^{kn_0} \frac{(kn_0)!}{N!(kn_0 - N)!} p_0^N (1 - p_0)^{kn_0 - N} \alpha_N = \alpha,$$

independent of the value of p_0. Since the left side of (6) is a polynomial in p_0, its constant term must equal α and all other coefficients must vanish. It will be observed that no terms of the sum in (6) that arise from $N > r$ will contribute to the coefficient of p_0^r; consequently this coefficient will not contain the unknowns $\alpha_{r+1}, \cdots, \alpha_{kn_0}$. These considerations show that the α_N must satisfy equations of the form

$$\alpha = c_{00}\,\alpha_0$$

$$0 = c_{10}\ \alpha_0 + c_{11}\ \alpha_1$$

$$\cdot \quad \cdot \quad\quad \cdot \quad\quad \cdot$$
$$\cdot \quad \cdot \quad\quad \cdot \quad\quad\quad \cdot$$
$$\cdot \quad \cdot \quad\quad \cdot \quad\quad\quad\quad \cdot$$

$$0 = c_{kn_00}\,\alpha_0 + c_{kn_01}\,\alpha_1 + \cdots + c_{kn_0kn_0}\,\alpha_{kn_0}\,.$$

It will also be observed that $c_{rr} = (kn_0)!/r!(kn_0 - r)!$; consequently the triangular matrix of the coefficients in these $kn_0 + 1$ non-homogeneous equations is non-singular. The equations therefore possess a unique solution, namely the known solution of $\alpha_N = \alpha$.

The preceding discussion shows that it is necessary to find critical regions of size α in each plane $\sum_1^k x_i = N$, $(N = 0, 1, \cdots, kn_0)$, if a critical region independent of p_0 is desired. If each such planar critical region were a best critical region for that plane, then the totality of such regions would constitute a best critical region independent of p_0 for testing H_0' against H_1'.

It follows from the theory of best critical regions [5] that if a best critical region in the plane $\sum_1^k x_i = N$ did exist, it would be determined by the inequality

$$(7) \qquad\qquad \frac{P_0[x_i \mid N]}{P_1[x_i \mid N]} < K,$$

where P_1 corresponds to P_0 when H_1 is true and where K is a constant whose value is chosen to make the critical region one of size α. Now from (4),

$$(8) \qquad \frac{P_0[x_i \mid N]}{P_1[x_i \mid N]} = \frac{(n_0!)^k (kn_0 - N)!\,(kn_1)!\,\Pi(n_1 - x_i)!}{(n_1!)^k (kn_1 - N)!\,(kn_0)!\,\Pi(n_0 - x_i)!}\,.$$

In order to study the possibility of a best critical region, it is therefore necessary to study the possibility of (8) satisfying inequality (7).

5. Approximate best critical region. Unfortunately, because the variables x_i are discrete, it is not possible to find critical regions of exactly size α for arbitrary α as required in (5). Consequently it is necessary to introduce continuous approximating functions for discrete probability functions or to resort to other devices if critical regions of the type discussed in the preceding section are to be obtained.

For the purpose of introducing such approximations, (8) will be written in the following form:

$$(9) \qquad \frac{P_0[x_i \mid N]}{P_1[x_i \mid N]} = c_1 \frac{(kn_0 - N)!}{\Pi(n_0 - x_i)!}\left(\frac{1}{k}\right)^{kn_0 - N} \div \frac{(kn_1 - N)!}{\Pi(n_1 - x_i)!}\left(\frac{1}{k}\right)^{kn_1 - N},$$

where c_1 is independent of the variables x_i. It will be observed that the ratio on the right is a ratio of two multinomial functions. Now the multinomial function

$$\frac{N!}{x_1!\,x_2!\cdots x_k!}\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

where $\sum_1^k x_i = N$, can be approximated by the multivariate normal function

$$(10) \qquad \frac{e \exp\left[-\tfrac{1}{2} \sum_1^k \left(\frac{x_i - Np_i}{\sqrt{Np_i}} \right)^2 \right]}{(2\pi N)^{\frac{1}{2}(k-1)} \sqrt{p_1 p_2 \cdots p_k}}.$$

The approximation is good provided the Np_i are large and the x_i remain away from their extreme values. If this approximation is applied to both numerator and denominator of (9), to this order of approximation,

$$(11) \qquad \frac{P_0[x_i \mid N]}{P_1[x_i \mid N]} = c_1 \frac{k^{k/2}\, e \exp\left[-\tfrac{1}{2} \sum_1^k \left(\frac{x_i - N/k}{\sqrt{n_0 - N/k}} \right)^2 \right]}{[2\pi(kn_0 - N)]^{\frac{1}{2}(k-1)}}$$

$$\div \frac{k^{k/2}\, e \exp\left[-\tfrac{1}{2} \sum_1^k \left(\frac{x_i - N/k}{\sqrt{n_1 - N/k}} \right)^2 \right]}{[2\pi(kn_1 - N)]^{\frac{1}{2}(k-1)}}$$

$$= c_1 \left[\frac{kn_1 - N}{kn_0 - N} \right]^{\frac{1}{2}(k-1)} c \exp\left[-\tfrac{1}{2} \frac{n_1 - n_0}{(n_1 - N/k)(n_0 - N/k)} \right.$$

$$\left. \cdot \sum_1^k (x_i - N/k)^2 \right].$$

Since, by hypothesis, $n_1 > n_0$ and $n_0 > N/k$, except for the case of $n_0 = N/k$, which will be considered later, it follows that

$$\frac{n_1 - n_0}{(n_1 - N/k)(n_0 - N/k)} > 0.$$

As a consequence, the right side of (11) will decrease in value as $\sum_1^k (x_i - N/k)^2$ increases in value. If $(x_1, \cdots, x_k)$ is a point lying on the sphere

$$(12) \qquad \sum_1^k (x_i - N/k)^2 = R$$

and if the coordinates of this point satisfy inequality (7) when approximation (11) is used, then all points outside this sphere will also satisfy (7) to this same order of approximation. A best critical planar region of size α in this approximate sense can therefore be obtained in the plane $\sum_1^k x_i = N$ by determining a

sphere with center at $(N/k, \cdots, N/k)$ such that when H_0' is true the probability is α that a point lying in the plane will lie outside this sphere. Furthermore, such a region will be a common best critical region for all values of $n_1 > n_0$ because the preceding arguments do not require the value of n_1 but merely the knowledge that $n_1 > n_0$.

For the purpose of determining the radius of the sphere that will yield the desired critical region, (4) will be expressed as follows:

$$(13) \qquad P_0[x_i \mid N] = c_2 \frac{N!}{\Pi x_i!} \left(\frac{1}{k}\right)^N \frac{(kn_0 - N)!}{\Pi(n_0 - x_i)!} \left(\frac{1}{k}\right)^{kn_0 - N},$$

where c_2 is independent of the x_i. If these multinomials are replaced by their multivariate normal approximations as given by (10), to this approximation (13) will reduce to

$$P_0[x_i \mid N] = c_3 \, e \, \exp\left[-\tfrac{1}{2} \sum_1^k \left(\frac{x_i - N/k}{\sqrt{N/k}}\right)^2 \right] e \exp\left[-\tfrac{1}{2} \sum_1^k \left(\frac{x_i - N/k}{\sqrt{n_0 - N/k}}\right)^2 \right]$$

$$(14)$$

$$= c_3 \, e \, \exp\left[-\tfrac{1}{2} \frac{\sum_1^k (x_i - N/k)^2}{\frac{N}{k}\left(1 - \frac{N}{kn_0}\right)} \right]$$

where c_3 is independent of the x_i. Since $\sum_1^k x_i = N$ here, x_k may be expressed in terms of the remaining variables; consequently (14), except for a constant factor, may be treated as a normal distribution in the variables $x_1, \cdots, x_{k-1}$. If the factorials in c_3 are replaced by their Stirling approximations, it will be found that c_3 is the correct constant for the normal distribution.

Since it is known [6] that -2 times the exponent in a normal distribution function possesses a chi-square distribution, it follows that to this order of approximation

$$(15) \qquad \frac{\sum_1^k (x_i - N/k)^2}{\frac{N}{k}\left(1 - \frac{N}{kn_0}\right)}$$

possesses a chi-square distribution with $k - 1$ degrees of freedom. If χ_α^2 is a value such that $P[\chi^2 > \chi_\alpha^2] = \alpha$, then

$$(16) \qquad \frac{\sum_1^k (x_i - N/k)^2}{\frac{N}{k}\left(1 - \frac{N}{kn_0}\right)} = \chi_\alpha^2$$

determines a sphere such that to this order of approximation the probability is α that a point lying in the plane $\sum_1^k x_i = N$ will lie outside the sphere. From

the arguments following (12), it therefore follows that a common best critical region in this approximate sense for testing H_0' against H_1' will consist of that part of each plane $\sum_{1}^{k} x_i = N$, $(N = 0, 1, \cdots, kn_0)$, which lies outside the corresponding sphere given by (16). Since the x_i are non-negative and do not exceed n_0, the planes corresponding to $N = 0$ and $N = kn_0$ contain a single point; therefore it is necessary to adopt some convention that assigns 100α percent of the samples with $N = 0$ and $N = kn_0$ to a critical region in order to obtain critical regions of size α in these two cases.

For a given set of data, the procedure to be followed then consists in calculating the statistic

$$ z = \frac{\sum_{1}^{k} (x_i - \bar{x})^2}{\bar{x}\left(1 - \dfrac{\bar{x}}{n_0}\right)}, $$

where $\bar{x} = \sum_{1}^{k} x_i/k$, and agreeing to reject the hypothesis that $n = n_0$ in favor of the alternative hypothesis that $n > n_0$ if and only if $z > \chi_\alpha^2$, where $P[\chi^2 > \chi_\alpha^2] = \alpha$ for $k - 1$ degrees of freedom. Because of the nature of the approximations used in (10) and (14), this result may be expected to be accurate only if $\bar{x}$ and $n_0 - \bar{x}$ are large.

The interesting feature of this result is that the familiar binomial index of dispersion, z, possesses optimum properties in this approximate sense for testing $n = n_0$ against $n > n_0$.

6. Poisson application. Since the preceding test will possess approximate optimum properties for n as large as desired, independent of the value of p, and since a Poisson distribution with parameter m can be approximated as closely as desired by means of a binomial distribution with $np = m$ by allowing n to increase sufficiently, it follows that the test will also possess approximate optimum properties for deciding between a binomial distribution with $n = n_0$ and a Poisson distribution.

7. Estimation of n. Although the purpose of this paper has been accomplished in the preceding sections, it is interesting to observe the role played by the closely related Poisson index of dispersion in the estimation of n.

Approximate confidence limits for n may be obtained by means of (16). If $\chi_{1-\alpha}^2$ is a value of χ^2 such that $P[\chi^2 > \chi_{1-\alpha}^2] = 1 - \alpha$, then, to this same order of approximation, the probability is $1 - 2\alpha$ that

$$ \chi_{1-\alpha}^2 < \frac{\sum_{1}^{k} (x_i - \bar{x})^2}{\bar{x}\left(1 - \dfrac{\bar{x}}{n}\right)} < \chi_\alpha^2. $$

If these inequalities are solved for n, the following $100(1 - 2\alpha)$ percent approximate confidence limits for n will be obtained:

$$(17) \qquad \frac{\bar{x}\chi_\alpha^2}{\chi_\alpha^2 - \dfrac{\Sigma(x_i - \bar{x})^2}{\bar{x}}} < n < \frac{\bar{x}\chi_{1-\alpha}^2}{\chi_{1-\alpha}^2 - \dfrac{\Sigma(x_i - \bar{x})^2}{\bar{x}}}.$$

Only the lower limit here will possess optimum properties. Now it will be observed that only positive values of n will be admissible if

$$\frac{\Sigma(x_i - \bar{x})^2}{\bar{x}} \leq \chi_{1-\alpha}^2,$$

whereas only negative values will be admissible if

$$\frac{\Sigma(x_i - \bar{x})^2}{\bar{x}} \geq \chi_\alpha^2.$$

The range of values will be infinite in each case if there is equality rather than inequality. If, however,

$$\chi_{1-\alpha}^2 < \frac{\Sigma(x_i - \bar{x})^2}{\bar{x}} < \chi_\alpha^2,$$

then both positive and negative values of n over infinite ranges will be admissible. Since n increases as the Poisson index $\Sigma(x_i - \bar{x})^2/\bar{x}$ increases until it becomes infinite and then increases from minus infinity through negative values, (17) may still be thought of as giving an interval (infinite) of values with a positive "lower" limit and a negative "upper" limit. Thus, the familiar Poisson index of dispersion plays an interesting role in determining whether a Poisson assumption is reasonable as far as admissible values of n are concerned.

If the population is truly binomial, negative values of n must be ruled out; consequently a Poisson assumption becomes increasingly tenable as the Poisson index increases. However, experience has shown [7] that a negative binomial distribution is often more realistic in describing data supposedly drawn from a binomial 'or Poisson population than is the assumed distribution; consequently a negative binomial should be given consideration if (17) yields only negative values or if it yields a negative "upper" limit that is numerically small relative to a positive "lower" limit.

It is also interesting to consider the point estimation of n. Here, it is customary [7] to estimate n by means of

$$\frac{k\bar{x}}{k - \dfrac{\Sigma(x_i - \bar{x})^2}{\bar{x}}}.$$

Thus, a positive, infinite, or negative estimate for n will be obtained according as the Poisson index is less than, equal to, or greater than k.

REFERENCES

[1] J. BERKSON, "Some difficulties of interpretation encountered in the application of the chi-square test," *Jour. Amer. Stat. Assoc.*, Vol. 33 (1938), pp. 526–536.

[2] B. H. CAMP, "Further interpretations of the chi-square test," *Jour. Amer. Stat. Assoc.*, Vol. 33 (1938), pp. 537–542.

[3] J. BERKSON, "A note on the chi-square test, the Poisson, and the binomial," *Jour. Amer. Stat. Assoc.*, Vol. 35 (1940), pp. 362–367.

[4] P. G. HOEL, "Testing the homogeneity of Poisson frequencies," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 362–368.

[5] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. Phil. Trans.*, Vol. 231 (1933), pp. 289–337.

[6] S. S. WILKS, *Mathematical Statistics*, Princeton Univ. Press, 1943, p. 104.

[7] "STUDENT," "An explanation of deviations from Poisson's law in practice," *Biometrika*, Vol. 12 (1919), pp. 211–215.

BILINEAR FORMS IN NORMALLY CORRELATED VARIABLES

By Allen T. Craig

University of Iowa

1. Summary. If a variable x is normally distributed with mean zero, we have previously given a necessary and sufficient condition (see references at end of this paper) for the independence of two real symmetric quadratic forms in n independent values of that variable. This condition is that the product of the matrices of the forms should vanish. In the present paper, we have proved that the same algebraic condition is both necessary and sufficient for the independence of two real symmetric bilinear, or a real symmetric bilinear and quadratic form, in normally correlated variables.

2. Introduction. In this paper, we determine the moment generating function of the joint distribution of two real symmetric bilinear forms in certain normally correlated variables and derive a necessary and sufficient condition for the independence, in the probability sense, of these forms. We further investigate the condition for independence, in the probability sense, of real symmetric bilinear and quadratic forms.

3. The moment generating function of the distribution of real symmetric bilinear forms. Let the two variables x and y have a joint normal distribution with means zero, unit variances and correlation coefficient ρ. From this bivariate distribution, repeated random samples of n pairs, say (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n), are drawn. Let $C = \| c_{jk} \|$ be a real symmetric matrix and write $\theta = \Sigma\Sigma c_{jk}x_jy_k$. The moment generating function of the distribution of θ is then given by

$$\varphi(t) = E[e^{t\theta}] = \frac{1}{(2\pi\sqrt{1-\rho^2})^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t\theta-Q} \, dy_n \, dx_n \cdots dy_1 \, dx_1,$$

where

$$Q = \frac{1}{2(1-\rho^2)} \sum_j (x_j^2 + y_j^2 - 2\rho x_j y_j)$$

and θ is defined above. If we subject the x's and y's to the same linear homogeneous transformation with appropriately chosen orthogonal matrix L, then Q remains invariant and θ becomes $\Sigma_j\lambda_j x_j'y_j'$ where the λ's are the n real roots of the characteristic equation of C, that is, of $| C - \lambda I | = 0$. The integrations are then easily effected and we find that

$$\varphi(t) = \left\{ \prod_j [1 - t(\rho + 1)\lambda_j][1 - t(\rho - 1)\lambda_j] \right\}^{-\frac{1}{2}},$$

$$= \left\{ | I - t(\rho + 1)C | \cdot | I - t(\rho - 1)C | \right\}^{-\frac{1}{2}}$$

$$= | I - 2\rho tC - (1 - \rho^2)t^2 C^2 |^{-\frac{1}{2}},$$

where I is the unit matrix of order n and the vertical bars, as usual, indicate the determinant of the enclosed matrix.

Next, let $A = \| a_{jk} \|$ and $B = \| b_{jk} \|$ be two real symmetric matrices each of order n. Write $\theta_1 = \Sigma\Sigma a_{jk}x_jy_k$ and $\theta_2 = \Sigma\Sigma b_{jk}x_jy_k$ where the x's and y's are the items of the sample randomly drawn from the bivariate distribution previously described. The moment generating function of the joint distribution of θ_1 and θ_2 is then given by

$$\varphi(t_1, t_2) = E[e^{t_1\theta_1 + t_2\theta_2}]$$

$$= (2\pi\sqrt{1-\rho^2})^{-n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t_1\theta_1 + t_2\theta_2 - Q} \, dy_n \, dx_n \cdots dy_1 \, dx_1,$$

where θ_1, θ_2, and Q have the meanings previously assigned to them. If we pursue a line of reasoning similar to that above, we find that

$$\varphi(t_1, t_2) = \left| I - 2\rho(t_1A + t_2B) - (1 - \rho^2)(t_1A + t_2B)^2 \right|^{-\frac{1}{2}}.$$

4. The independence of bilinear forms. It is clear that there exist positive numbers, say h_1 and h_2, such that $\varphi(t_1, t_2)$ exists for $0 < t_1 < h_1$ and $0 < t_2 < h_2$. It is well known that a necessary and sufficient condition for the independence of θ_1 and θ_2 is that $\varphi(t_1, t_2)$ shall factor into the product $\varphi(t_1, 0)\varphi(0, t_2)$. If then, we assume θ_1 and θ_2 to be independent, we have essentially

(1)
$$\left| I - 2\rho(t_1A + t_2B) - (1 - \rho^2)(t_1A + t_2B)^2 \right|$$
$$= \left| I - 2\rho t_1A - (1 - \rho^2)t_1^2A^2 \right| \cdot \left| I - 2\rho t_2B - (1 - \rho^2)t_2^2B^2 \right|.$$

If h denotes the smaller of h_1 and h_2, then the factored form holds for $0 < t_1, t_2 < h$, and hence for all real values of t_1 and t_2. In particular it holds for $t_2 = t_1$ so that

$$\left| I - 2\rho t_1(A + B) - (1 - \rho^2)t_1^2(A + B)^2 \right|$$
$$= \left| I - 2\rho t_1A - (1 - \rho^2)t_1^2A^2 \right| \cdot \left| I - 2\rho t_1B - (1 - \rho^2)t_1^2B^2 \right|.$$

Let r_1, r_2, and $r \leq r_1 + r_2$ denote the ranks of the matrices A, B, and $A + B$. Further let the real non-zero roots of the characteristic equations of these matrices be denoted respectively by $\alpha_1, \alpha_2, \cdots, \alpha_{r_1}, \beta_1, \beta_2, \cdots, \beta_{r_2}$, and $\gamma_1, \gamma_2, \cdots, \gamma_r$. Then the members of the preceding equation may be written

$$\prod_{i=1}^{r} [1 - t_1(\rho + 1)\gamma_i][1 - t_1(\rho - 1)\gamma_i]$$

and

$$\prod_{i=1}^{r_1} [1 - t_1(\rho + 1)\alpha_i][1 - t_1(\rho - 1)\alpha_i] \prod_{i=1}^{r_2} [1 - t_1(\rho + 1)\beta_i][1 - t_1(\rho - 1)\beta_i]$$

respectively. It is seen that the left member is a polynomial in t_1 of degree $2r$ and that the right member is a polynomial in t_1 of degree $2(r_1 + r_2)$. Accord-

ingly, $r = r_1 + r_2$ and the roots $\gamma_1, \cdots, \gamma_1$ consist of the roots $\alpha_1, \cdots, \alpha_{r_1}$, $\beta_1, \cdots, \beta_{r_2}$. That is, if θ_1 and θ_2 are independent, then the rank of $A + B$ is the sum of the ranks of A and B and the non-zero roots of the characteristic equation of $A + B$ consist of those of the characteristic equation of A together with those of B. Further, if in (1) we put $t_2 = vt_1$, where v is real, we have

$$| I - 2\rho t_1(A + vB) - (1 - \rho^2)t_1^2(A + vB)^2 |$$

$$= | I - 2\rho t_1 A - (1 - \rho^2)t_1^2 A^2 | \cdot | I - 2\rho t_1 vB - (1 - \rho^2)t_1^2 v^2 B^2 |.$$

Denote the rank of $A + vB$ by r' and the non-zero roots of its characteristic equation by $\delta_1, \cdots, \delta_{r'}$. The immediately preceding equation can then be written

$$\prod_{i=1}^{r'} [1 - t_1(\rho + 1)\delta_i][1 - t_1(\rho - 1)\delta_i]$$

$$= \prod_{i=1}^{r_1} [1 - t_1(\rho + 1)\alpha_i][1 - t_1(\rho - 1)\alpha_i] \prod_{i=1}^{r_2} [1 - t_1(\rho + 1)v\beta_i][1 - t_1(\rho - 1)v\beta_i].$$

From this we infer that, apart from zero roots, the roots of the characteristic equation of $A + vB$ are $\alpha_1, \cdots, \alpha_{r_1}, v\beta_1, \cdots, v\beta_{r_2}$.

If a symmetric matrix, say $M(v)$, has elements which are real polynomials in the real variable v, and if the determinant

$$| M(v) - \lambda I | = (-1)^n[\lambda - p_1(v)][\lambda - p_2(v)] \cdots [\lambda - p_n(v)],$$

where $p_1(v), p_2(v), \cdots, p_n(v)$ are likewise real polynomials in v, then there exists, for all real values of v, a real orthogonal matrix, say $L(v)$, such that

$$L'(v)M(v)L(v) = \begin{Vmatrix} p_1(v) & 0 & \cdots & 0 \\ 0 & p_2(v) & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & & & p_n(v) \end{Vmatrix}.$$

Furthermore[1], $\dfrac{dL(v)}{dv}$ exists for all real values of v. Since

$$| A + vB - \lambda I | = (-1)^n \lambda^{n-(r_1+r_2)}(\lambda - \alpha_1) \cdots (\lambda - \alpha_{r_1})(\lambda - v\beta_1) \cdots (\lambda - v\beta_{r_2}),$$

then $A + vB$ belongs to the class $M(v)$ so we have

$$(2) \qquad L'(v)(A + vB)L(v) = \begin{Vmatrix} \alpha_1 & 0 & \cdots & & 0 \\ \cdot & \cdot & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ 0 & \cdots & \alpha_{r_1} & \cdots & 0 \\ 0 & \cdots & v\beta_1 & \cdots & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & & \cdots & v\beta_{r_2} & \cdots & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & & & & 0 \end{Vmatrix}.$$

In particular,

$$(3) \qquad L'(0)AL(0) = \begin{Vmatrix} \alpha_1 & \cdots & & & 0 \\ \cdot & \cdot & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ & & \alpha_{r_1} & & \\ & & & 0 & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & & & & 0 \end{Vmatrix}.$$

If we differentiate (2) with respect to v and subsequently set $v = 0$, we have

$$(4) \qquad \frac{dL'(0)}{dv} AL(0) + L'(0)BL(0) + L'(0)A\frac{dL(0)}{dv} = \begin{Vmatrix} 0 & \cdots & & & 0 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & & \cdot & & \\ & & 0 & & \\ & & \beta_1 & & \\ & & & \cdot & \\ & & & \cdot & \\ & & & \cdot & \\ & & & \beta_{r_2} & \\ & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & & & & 0 \end{Vmatrix}.$$

Since $L(v)$ is orthogonal, then $L'(v)L(v) = I$. Upon differentiating both members with respect to v, and subsequently setting $v = 0$, it is seen that $\dfrac{dL'(0)}{dv} L(0) = -L'(0) \dfrac{dL(0)}{dv}$ so that $L'(0) \dfrac{dL(0)}{dv}$ is a skew-symmetric matrix, say S. Further

$$(5) \qquad \frac{dL'(0)}{dv} = -L'(0) \frac{dL(0)}{dv} L'(0) = -SL'(0),$$

and, by taking conjugates,

$$(6) \qquad \frac{dL(0)}{dv} = -L(0) \frac{dL'(0)}{dv} L(0) = L(0)S.$$

If we multiply (5) on the right by $AL(0)$ and (6) on the left by $L'(0)A$, we see that (4) may be written

$$(7) \qquad L'(0)BL(0) = \begin{Vmatrix} 0 & \cdots & & & 0 \\ & & & & \cdot \\ & & & & \cdot \\ & & 0 & & \cdot \\ & & \beta_1 & & \cdot \\ & & & & \cdot \\ & & & \cdot & \cdot \\ & & & & \cdot \\ & & & \beta_{r_2} & \cdot \\ & & & 0 & \cdot \\ & \cdot & & \cdot & \cdot \\ & \cdot & & \cdot & \\ 0 & & & 0 & \end{Vmatrix} + SL'(0)AL(0) - L'(0)AL(0)S.$$

Since S is skew-symmetric and since $L'(0)AL(0)$ is given by (3), then each element on the principal diagonal of $SL'(0)AL(0)$ and $L'(0)AL(0)S$ is zero. Further, since $L'(0)BL(0)$ is symmetric, then $L'(0)BL(0)$ takes the form

$$\begin{Vmatrix} 0 & k_{12} & \cdots & & & k_{1n} \\ k_{12} & 0 & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \beta_1 & \\ & & & & & \cdot \\ & & & & & \cdot \\ & & & & \beta_{r_2} & \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ k_{1n} & & & & & 0 \end{Vmatrix}$$

Because the non-zero roots of the characteristic equation of $L'(0)BL(0)$ are β_1, $\cdots$, β_{r_2} then the sum of all two-rowed principal minors of the determinant of $L'(0)BL(0)$ must equal the sum of the products of β_1, $\cdots$, β_{r_2} taken two at a time. That is

$$\sum_{i<j} \beta_i \beta_j = \sum_{i<j} \beta_i \beta_j - \Sigma\, k_{ij}^2,$$

so that each k_{ij}, being real, is zero. Accordingly, $SL'(0)AL(0) - L'(0)AL(0)S$ is a zero matrix and $L'(0)BL(0)$ is given by the first term in the right member of (7). We then have

$$L'(0)AL(0)L'(0)BL(0) = L'(0)ABL(0) = 0,$$

from which it follows that $AB = 0$. Thus, if the real symmetric bilinear forms θ_1 and θ_2 are independent in the probability sense, the product of their matrices is zero.

If, conversely, $AB = 0$, then

$$\varphi(t_1, t_2) = |\, I - 2\rho(t_1 A + t_2 B) - (1 - \rho^2)(t_1^2 A^2 + t_2^2 B^2)\,|^{-\frac{1}{2}},$$

$$= |\,[I - 2\rho t_1 A - (1 - \rho^2)t_1^2 A^2][I - 2\rho t_2 B - (1 - \rho^2)t_2^2 B^2]\,|^{-\frac{1}{2}},$$

$$= \varphi(t_1, 0)\varphi(0, t_2),$$

and θ_1 and θ_2 are independent. This establishes the following theorem.

THEOREM I. *Let x and y be normally correlated with means zero, unit variances, and correlation coefficient ρ. Let θ_1 and θ_2 be two real symmetric bilinear forms in n random pairs of values of x and y, say $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$. A necessary and sufficient condition that θ_1 and θ_2 be independent in the probability sense is that the product of the matrices of the forms be zero.*

5. Simultaneous reduction of quadratic or bilinear forms. The argument of Section 4 may be used to establish in a very simple manner the following theorem.

THEOREM II. *Let A and B be two real symmetric matrices with constant elements, each matrix of order n. A necessary and sufficient condition that there exist a real orthogonal matrix of order n such that simultaneously each of $L'AL$ and $L'BL$ is in canonical form, wherein no non-zero elements occupy corresponding positions on the principal diagonals, is that $AB = 0$.*

For if such an orthogonal matrix L exists, it is evident that $L'ALL'BL = L'ABL = 0$ from which it follows that $AB = 0$. Conversely, if $AB = 0$, then v being a real scalar, the matrix $(A - \lambda I)(vB - \lambda I)$ is equal to the matrix $-\lambda[(A + vB) - \lambda I]$. These matrices being equal, their determinants are equal so that $A + vB$ belongs to the class $M(v)$ of section 4. Thus L may be taken as $L(0)$ and simultaneously $L'AL$ and $L'BL$ are of the form stated in the theorem.

6. Independence of bilinaer and quadratic forms. Let $\theta = \Sigma\Sigma a_{jk} x_j y_k$ be a real symmetric bilinear form of rank r_1 in the previously defined variables

$(x_1, y_1), \cdots, (x_n, y_n)$ and let $q = \Sigma\Sigma b_{jk}x_jx_k$ be a real symmetric quadratic form of rank r_2 in $x_1, x_2, \cdots, x_n$. As usual, denote the non-zero roots of the characteristic equations of A and B by $\alpha_1, \alpha_2, \cdots, \alpha_{r_1}$ and $\beta_1, \beta_2, \cdots, \beta_{r_2}$ respectively. The moment generating function of the joint distribution of θ and q is

$$\varphi(t_1, t_2) = \frac{1}{(2\pi\sqrt{1-\rho^2})^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t_1\theta + t_2 q - Q}\, dy_n\, dx_n \cdots dy_1\, dx_1,$$

where, as previously,

$$Q = \frac{1}{2(1-\rho^2)} \Sigma(x_j^2 + y_j^2 - 2\rho x_j y_j).$$

We first orthogonally transform the variables so that the exponent in the integrand becomes, upon writing $||f_{jk}|| = L'BL$,

$$t_1 \Sigma\alpha_j x_j' y_j' + t_2 \Sigma\Sigma f_{jk} x_j' x_k' - \frac{1}{2(1-\rho^2)} \Sigma(x_j'^2 + y_j'^2 - 2\rho x_j' y_j').$$

We then integrate on $y_1', y_2', \cdots, y_n'$ and obtain for the exponent in the integrand

$$t_2 \Sigma\Sigma f_{jk} x_j' x_k' - \tfrac{1}{2}\Sigma x_j'^2 + \rho t_1 \Sigma\alpha_j x_j'^2 + \frac{1-\rho^2}{2} t_1^2 \Sigma\alpha_j^2 x_j'^2 .$$

If we effect on the variables $x_1', x_2', \cdots, x_n'$ the inverse of the orthogonal transformation initially used on the x's and y's, the exponent in the integrand becomes, using $||g_{jk}|| = A^2$,

$$t_2 \Sigma\Sigma b_{jk} x_j x_k - \tfrac{1}{2}\Sigma x_j^2 + \rho t_1 \Sigma\Sigma a_{jk} x_j x_k + \frac{1-\rho^2}{2} t_1^2 \Sigma\Sigma g_{jk} x_j x_k$$

or

$$- \tfrac{1}{2}\Sigma\Sigma[\delta_{jk} - 2\rho t_1 a_{jk} - (1-\rho^2)t_1^2 g_{jk} - 2t_2 b_{jk}]x_j x_k ,$$

where δ_{jk} equals 1 or 0 according as j does or does not equal k. Hence,

(8) $$\varphi(t_1, t_2) = |I - 2\rho t_1 A - (1-\rho^2)t_1^2 A^2 - 2t_2 B|^{-\frac{1}{2}}.$$

If θ and q are independent, we have

(9) $$|I - 2\rho t_1 A - (1-\rho^2)t_1^2 A^2 - 2t_2 B|$$
$$= |I - 2\rho t_1 A - (1-\rho^2)t_1^2 A^2| \cdot |I - 2t_2 B|,$$

for $0 < t_1 < h_1$ and $0 < t_2 < h_2$. As before, the members of (9) are polynomials which, being equal for $0 < t_1, t_2 < h$, are equal for all real values of t_1 and t_2. If we put $t_1 = 1$ and $t_2 = vt_1 = v$, where v is real, then (9) becomes

$$|I - 2\rho A - (1-\rho^2)A^2 - 2vB| = |I - 2\rho A - (1-\rho^2)A^2| \cdot |I - 2vB|$$
$$= \prod_1^{r_1} [1 - (\rho-1)\alpha_j][1 - (\rho+1)\alpha_j]\prod_1^{r_2}(1 - 2v\beta_j).$$

That is,

$$| 2\rho A + (1 - \rho^2)A^2 + 2vB - \lambda I |$$

$$= (-1)^n \lambda^{n-(r_1+r_2)} [\lambda - 2\rho\alpha_1 - (1 - \rho^2)\alpha_1^2] \cdots [\lambda - 2\rho\alpha_{r_1} - (1 - \rho^2)\alpha_{r_1}^2]$$

$$\cdot [\lambda - 2v\beta_1] \cdots [\lambda - 2v\beta_{r_2}]$$

so that $2\rho A + (1 - \rho^2)A^2 + 2vB$ is a matrix of the class $M(v)$. Hence we write

$$L'(v)[2\rho A + (1 - \rho^2)A^2 + 2vB]L(v) = \begin{Vmatrix} 2\rho\alpha_1 + (1-\rho^2)\alpha_1^2 & \cdots & & & 0 \\ & \ddots & & & \\ & & 2\rho\alpha_{r_1} + (1-\rho^2)\alpha_{r_1}^2 & & \\ & & 2v\beta_1 & & \\ & & & \ddots & \\ & & & 2v\beta_{r_2} & \\ & & & & 0 \\ 0 & & & & 0 \end{Vmatrix}$$

The argument of section 4 shows that $L'(0)[2\rho A + (1 - \rho^2)A^2]L(0)L'(0)2BL(0)$ is a zero matrix, from which it follows that $2\rho AB + (1 - \rho^2) A^2B = 0$. But this imposes on ρ, n^2 conditions of the form

$$2\rho l_{jk} + (1 - \rho^2)m_{jk} = 0, \qquad (j, k = 1, 2, \cdots, n).$$

Since these hold for every $-1 < \rho < 1$, they hold identically. Hence each l_{jk} and m_{jk} is zero. In particular, $\| l_{jk} \| = AB = 0$ if θ and q are independent.

Conversely, if $AB = 0$, we see by Theorem II that (8) becomes

$$\varphi(t_1, t_2) = \varphi(t_1, 0)\varphi(0, t_2),$$

so that θ and q are independent. This yields Theorem III.

THEOREM III. *Let x and y be normally correlated with means zero, unit variances, and correlation coefficient ρ. Let θ be a real symmetric bilinear form in the n random pairs of values of x and y, say $(x_1, y_1), \cdots, (x_n, y_n)$, and let q be a real symmetric quadratic form in $x_1, x_2, \cdots, x_n$ (or $y_1, \cdots, y_n$). A necessary and sufficient condition that θ and q be independent in the probability sense is that the product of the matrices of the forms be zero.*

For example, let θ be n times the sample covariance and let q be n times the square of the mean of the x's. Then

$$\theta = \Sigma(x_j - \bar{x})(y_j - \bar{y})$$

$$= \Sigma\Sigma a_{jk}x_jy_k;$$

where

$$a_{jk} = \frac{n-1}{n} \quad \text{if} \quad j = k,$$

$$= -\frac{1}{n} \text{ otherwise,}$$

and

$$q = n\bar{x}^2 = \Sigma\Sigma b_{jk}x_jx_k, \qquad b_{jk} = 1/n \text{ for } j, k = 1, 2, \cdots, n.$$

Since $AB = 0$, then θ and q are independent, a fact otherwise known but perhaps not so easily established.

REFERENCES

[1] ALLEN T. CRAIG, "Note on the independence of certain quadratic forms", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 195–197.

[2] HAROLD HOTELLING, "Note on a matric theorem of A. T. Craig", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 427–429.

ON THE CHARLIER TYPE B SERIES

By S. Kullback

George Washington University

1. Introduction. The Type B series of Charlier has been discussed in some detail in the literature (See references at the end of the paper). The problem of the convergence of the Type B series has been considered by Pollaczek-Geiringer [12], [13], Szegö [12] (page 110), Uspensky [16], Jacob [5], Schmidt [16] and Obrechkoff [11]. There is presented in the following a method of development of the Type B series which is believed to be of some interest, including a necessary and sufficient condition for the convergence which is basically the same as that of Schmidt [16]. A result of Steffensen [17] is extended and shown to be related to the Charlier Type B series.

2. Statement of results. Consider the function $p(r)$, defined for $r = 0, 1, 2, \cdots$, and such that

$$(2.1) \qquad \sum_{r=0}^{\infty} p(r) = 1; \qquad \sum_{r=0}^{\infty} |p(r)| = A$$

where A is some finite value. Let the n-th factorial moment be defined by

$$(2.2) \qquad \begin{aligned} \mu_{(0)} &= 1 \\ \mu_{(n)} &= \sum_{r=0}^{\infty} r(r-1)(r-2)\cdots(r-n+1)p(r), \qquad (n = 1, 2, \cdots) \end{aligned}$$

For arbitrary λ let

$$(2.3) \qquad \begin{aligned} L_n &= \mu_{(n)} - n\mu_{(n-1)}\lambda + \frac{n(n-1)}{2!}\mu_{(n-2)}\lambda^2 \\ &\quad - \frac{n(n-1)(n-2)}{3!}\mu_{(n-3)}\lambda^3 + \cdots + (-1)^n\lambda^n. \end{aligned}$$

We prove the following results:

THEOREM. *A necessary and sufficient condition that the function $p(r)$ of (2.1) may be expressed by the absolutely convergent series*

$$(2.4) \qquad p(r) = \frac{e^{-\lambda}\lambda^r}{r!} + L_1 \frac{\partial}{\partial \lambda}\frac{e^{-\lambda}\lambda^r}{r!} + \frac{L_2}{2!}\frac{\partial^2}{\partial \lambda^2}\frac{e^{-\lambda}\lambda^r}{r!} + \cdots$$

is that

$$(2.5) \qquad 1 + |\mu_{(1)}| + \tfrac{1}{2!}|\mu_{(2)}| + \tfrac{1}{3!}|\mu_{(3)}| + \cdots + \frac{1}{n!}|\mu_{(n)}| + \cdots$$

converges where L_n is defined as in (2.3).

3. Generating functions. For the function $p(r)$ of (2.1) consider the generating function defined by

$$(3.1) \qquad \varphi(z) = \sum_{r=0}^{\infty} z^r p(r)$$

where z is a complex variable. Because of (2.1) it is clear that the right member of (3.1) is uniformly and absolutely convergent for $|z| \leq 1$ so that the radius of convergence of (3.1) is some value $R_1 > 1$.

The Taylor expansion of $\varphi(z)$ about the point $z = 1$ is given by

$$(3.2) \qquad \varphi(z) = \varphi(1) + (z - 1)\varphi'(1) + \frac{(z - 1)^2}{2!} \varphi''(1) + \cdots$$

where, as may be readily obtained from (3.1),

$$(3.3) \qquad \varphi^{(n)}(1) = \sum_{r=0}^{\infty} r(r - 1)(r - 2)\cdots(r - n + 1)p(r) = \mu_{(n)} .$$

If it is assumed that (2.5) converges, then

$$(3.4) \qquad \varphi(z) = 1 + (z - 1)\mu_{(1)} + \frac{(z - 1)^2}{2!} \mu_{(2)} + \cdots + \frac{(z - 1)^n}{n!} \mu_{(n)} + \cdots$$

is uniformly and absolutely convergent for $|z - 1| \leq 1$.

4. Sufficiency. For arbitrary λ let us set

$$(4.1) \qquad e^{-\lambda(z-1)} \left(1 + \mu_{(1)}(z - 1) + \mu_{(2)} \frac{(z - 1)^2}{2!} + \cdots \right)$$
$$= 1 + L_1(z - 1) + \frac{L_2}{2!} (z - 1)^2 + \cdots$$

where the right member, because of (3.4) is absolutely convergent for $|z - 1| \leq 1$. The coefficients on the right side of (4.1) are given by

$$(4.2) \qquad L_n = \mu_{(n)} - n\mu_{(n-1)} \lambda + \frac{n(n - 1)}{2!} \mu_{(n-2)} \lambda^2 - \cdots + (-1)^n \lambda^n$$

and the factorial moments may also be expressed by

$$(4.3) \qquad \mu_{(n)} = L_n + nL_{n-1} \lambda + \frac{n(n - 1)}{2!} L_{n-2} \lambda^2 + \cdots + \lambda^n.$$

These relations are readily derived by expressing (4.1) symbolically as

$$(4.4) \qquad e^{-\lambda(z-1) + \mu(z-1)} = e^{L(z-1)}$$

where after expansion μ^n and L^n are to be replaced by $\mu_{(n)}$ and L_n respectively. (Cf. Jordan [7], p. 39). From (4.1) and (3.4) there is now derived

$$(4.5) \qquad \varphi(z) = e^{\lambda(z-1)} \left(1 + L_1(z - 1) + \frac{L_2}{2!} (z - 1)^2 + \cdots \right).$$

Since the right member of (4.5) is absolutely and uniformly convergent for $|z - 1| \leqq 1$ for arbitrary λ, it may be expressed as

$$(4.6) \qquad \varphi(z) = \left(1 + L_1 \frac{\partial}{\partial \lambda} + \frac{L_2}{2!} \frac{\partial^2}{\partial \lambda^2} + \cdots \right) e^{\lambda(z-1)}.$$

Since the radius of convergence of the right member of (4.6) is some value R_2 such that $|z - 1| < R_2 > 1$, it may be expressed as a power series about $z = 0$, or

$$(4.7) \qquad \varphi(z) = \left(1 + L_1 \frac{\partial}{\partial \lambda} + \frac{L_2}{2!} \frac{\partial^2}{\partial \lambda^2} + \cdots \right) e^{-\lambda} \left(1 + \lambda z + \frac{\lambda^2 z^2}{2!} + \cdots \right).$$

Recalling now the definition of $\varphi(z)$ as given in (3.1), there is obtained by equating coefficients of like powers of z in (3.1) and (4.7)

$$(4.8) \qquad p(r) = \left(1 + L_1 \frac{\partial}{\partial \lambda} + \frac{L_2}{2!} \frac{\partial^2}{\partial \lambda^2} + \cdots \right) \frac{e^{-\lambda} \lambda^r}{r!}.$$

Since it may be readily shown that

$$(4.9) \qquad \frac{\partial^n}{\partial \lambda^n} \frac{e^{-\lambda} \lambda^r}{r!} = (-1)^n \Delta^n \frac{e^{-\lambda} \lambda^r}{r!}$$

where

$$\Delta \frac{e^{-\lambda} \lambda^r}{r!} = \frac{e^{-\lambda} \lambda^r}{r!} - \frac{e^{-\lambda} \lambda^{r-1}}{(r-1)!}$$

and

$$\Delta^n \frac{e^{-\lambda} \lambda^r}{r!} = \Delta^{n-1} \frac{e^{-\lambda} \lambda^r}{r!} - \Delta^{n-1} \frac{e^{-\lambda} \lambda^{r-1}}{(r-1)!}$$

we may also write (4.8) as

$$(4.10) \qquad p(r) = \frac{e^{-\lambda} \lambda^r}{r!} - L_1 \Delta \frac{e^{-\lambda} \lambda^r}{r!} + \frac{L_2}{2!} \Delta^2 \frac{e^{-\lambda} \lambda^r}{r!} - \frac{L_3}{3!} \Delta^3 \frac{e^{-\lambda} \lambda^r}{r!} + \cdots.$$

5. Necessity. Assume that the function $p(r)$ of (2.1), for arbitrary λ, is given by the absolutely convergent series

$$(5.1) \qquad p(r) = \left(1 + L_1 \frac{\partial}{\partial \lambda} + \frac{L_2}{2!} \frac{\partial^2}{\partial \lambda^2} + \cdots \right) \frac{e^{-\lambda} \lambda^r}{r!}.$$

Since $e^{-\lambda} \lambda^r / r!$ is continuous with respect to λ, there follows, where z is a complex variable and $|z| \leqq 1$

$$\sum_{r=0}^{\infty} z^r p(r) = \sum_{r=0}^{\infty} \frac{z^r e^{-\lambda} \lambda^r}{r!} + L_1 \frac{\partial}{\partial \lambda} \sum_{r=0}^{\infty} \frac{z^r e^{-\lambda} \lambda^r}{r!} + \frac{L_2}{2!} \frac{\partial^2}{\partial \lambda^2} \sum_{r=0}^{\infty} \frac{z^r e^{-\lambda} \lambda^r}{r!} + \cdots$$

$$(5.2) \qquad = e^{\lambda(z-1)} \left(1 + L_1(z - 1) + \frac{L_2}{2!}(z - 1)^2 + \cdots \right)$$

$$= 1 + M_1(z - 1) + \frac{M_2}{2!}(z - 1)^2 + \frac{M_3}{3!}(z - 1)^3 +$$

where

(5.3) $$M_n = L_n + nL_{n-1}\lambda + \frac{n(n-1)}{2!}L_{n-2}\lambda^2 + \cdots + \lambda^n.$$

From (5.2) it follows that

(5.4) $$M_n = \mu_{(n)}$$

where $\mu_{(n)}$ is as defined in (3.3). Since (5.1) becomes for $r = 0, \lambda = 0$

(5.5) $$1 - \mu_{(1)} + \frac{1}{2!}\mu_{(2)} - \frac{1}{3!}\mu_{(3)} + \cdots$$

the assumed absolute convergence implies that

(5.6) $$1 + |\mu_{(1)}| + \frac{1}{2!}|\mu_{(2)}| + \frac{1}{3!}|\mu_{(3)}| + \cdots + \frac{1}{n!}|\mu_{(n)}| + \cdots$$

converges.

6. Remarks. Obrechkoff [11] shows that his result includes those of Pollaczek-Geiringer [12], Szegö [12] (p. 110) and Jacob [5]. His theorem states that if the function $p(r)$, $(r = 0, 1, 2, \cdots)$, satisfies the following conditions

(6.1) $$\sum_{r=1}^{\infty} 2^r r^A \mid p(r) \mid$$

is convergent for each finite number A, and

(6.2) $$\frac{(4\lambda)^n}{(n+1)!}\sum_{r=1}^{n}\frac{\mid p(r) \mid}{r}(e^{-\lambda}\lambda^r/r!)^{-1}$$

tends toward zero as n increases indefinitely then $p(r)$ may be expressed in a convergent Charlier Type B series.

Uspensky [18] shows that if

(6.3) $$\sum_{r=0}^{\infty} z^r p(r)$$

has a radius of convergence $R > 2$ then $p(r)$ may be expressed in a convergent Charlier Type B series.

Schmidt [16] shows that a necessary and sufficient condition for the convergence is that the function $\varphi(z)$ defined as in (3.1) (he does not explicitly impose the condition (2.1) on $p(r)$) be regular inside the two circles $|z| < 1$ and $|z - 1| < 1$ and with all its derivatives is continuous on the peripheries also. In the case that $p(r) \geqq 0$, the condition (2.5) is stronger, in fact in this case Schmidt [16] shows that a necessary and sufficient condition is that

$$\lim_{r=\infty} p(r)2^r r^k = 0$$

for all integral $k \geqq 0$. If $p(r) \geqq 0$, then Uspensky's condition is only just enough stronger than Schmidt's to keep it from being sufficient.

If (6.1) is satisfied, or if (6.3) is satisfied then (3.1) is absolutely convergent for $|z| \leqq 2$. Therefore, the point $z = 2$ is contained in the circle of convergence of (3.2) or (3.4) which implies that

$$1 + |\mu_{(1)}| + \frac{1}{2!}|\mu_{(2)}| + \cdots + \frac{1}{n!}|\mu_{(n)}| + \cdots$$

converges.

It is deemed worthy of special mention to point out, as both Schmidt and Uspensky have done, the striking fact that the necessary and sufficient condition for the validity of (2.4) is independent of λ. This arbitrariness of λ enables us to dispose of it so as to obtain better convergence. Indeed if we set $\lambda = \mu_{(1)}$ then as is evident from (4.2) $L_1 = 0$.

7. Special cases. It is of interest to note that (4.8) is the Taylor expansion if $p(r) = e^{-\mu}\mu^r/r!$, $(r = 0, 1, 2, \cdots)$, for then (4.2) becomes

$$(7.1) \qquad\qquad L_n = (\mu - \lambda)^n$$

since for the Poisson Exponential Distribution $e^{-\mu}\mu^r/r!$, $(r = 0, 1, 2, \cdots)$, $\mu_{(n)} = \mu^n$ and (4.8) is then

$$(7.2) \qquad \frac{e^{-\mu}\mu^r}{r!} = \frac{e^{-\lambda}\lambda^r}{r!} + (\mu - \lambda)\frac{\partial}{\partial\lambda}\frac{e^{-\lambda}\lambda^r}{r!} + \frac{(\mu - \lambda)^2}{2!}\frac{\partial^2}{\partial\lambda^2}\frac{e^{-\lambda}\lambda^r}{r!} + \cdots.$$

If $p(r)$ is finite, that is if $p(r) = 0$ for $r \geq n + 1$ then $\mu_{(k)} = 0$ for $k \geq n + 1$. Thus, for a finite function the condition (2.5) is satisfied.

8. Factorial moments. For functions $p(r)$, $(r = 0, 1, 2, \cdots)$, satisfying (2.5), there may be derived from (3.1) and (3.4) the relation

$$(8.1) \qquad r!p(r) = \mu_{(r)} - \mu_{(r+1)} + \frac{1}{2!}\mu_{(r+2)} - \frac{1}{3!}\mu_{(r+3)} + \cdots , \quad (r = 0, 1, 2, \cdots),$$

since each side is $\varphi^{(r)}(0)$ derived respectively from (3.1) and (3.4). It should be noted that for $\lambda = 0$ (4.5) leads to (8.1) rather than (4.8) so that (8.1) may be considered as the Charlier Type B series for $\lambda = 0$. The result (8.1) was derived for finite functions by Steffensen [17]. (Also compare Kaplansky [8]). This may also be expressed symbolically by

$$(8.2) \qquad\qquad p(r) = \mu^r e^{-\mu}/r!, \qquad\qquad (r = 0, 1, 2, \cdots),$$

where after expansion μ^n is to be replaced by $\mu_{(n)}$. It is of interest to note the relation between the symbolic expression for $p(r)$ as a Poisson Exponential in (8.2) and the series (4.8), for (4.8) may be expressed symbolically as

$$(8.3) \qquad\qquad p(r) = e^{L(\partial/\partial\lambda)} \cdot \frac{e^{-\lambda}\lambda^r}{r!} = e^{-(\lambda+L)}(\lambda + L)^r/r!$$

$$= \mu^r e^{-\mu}/r!$$

since $e^{a(d/dx)}f(x) = f(x + a)$ and the relations (4.2), (4.3), (4.4).

9. Illustrations. Consider the function

(9.1) $$p(r) = 1/2^{r+1},$$ $(r = 0, 1, 2, \cdots).$

For this function

(9.2) $$\varphi(z) = \sum_{r=0}^{\infty} z^r p(r) = 1/(2 - z)$$

and

(9.3) $$\varphi^{(n)}(1) = \mu_{(n)} = n!$$

so that (2.5) becomes

(9.4) $$1 + 1 + 1 + \cdots$$

which does not converge. (It may be of interest to note that for this case (8.1) yields

(9.5) $$p(0) = 1 - 1 + 1 - 1 + 1 - \cdots .$$

The series on the right in (9.5) is not convergent but is summable C_1 to $\frac{1}{2}$. For the latter see for example R. P. Agnew, [19].) In this case the first several coefficients of (4.8) are for $\lambda = 1$,

(9.6)
$$L_1 = 0, \qquad \frac{L_2}{2!} = .5000, \qquad \frac{L_3}{3!} = .3333, \qquad \frac{L_4}{4!} = .3750$$
$$\frac{L_5}{5!} = .3667, \qquad \frac{L_6}{6!} = .3681, \qquad \frac{L_7}{7!} = .3679, \qquad \cdots .$$

Let us now consider the function

(9.7) $$p(0) = \tfrac{1}{2}, \qquad p(r) = \tfrac{1}{3}^r,$$ $(r = 1, 2, \cdots).$

For this function

(9.8) $$\varphi(z) = \sum_{r=0}^{\infty} z^r p(r) = \tfrac{1}{2} + \frac{z}{3 - z}$$

and

(9.9) $$\varphi^{(n)}(1) = \mu_{(n)} = \frac{n!}{2^n}\binom{3}{2},$$ $(n = 1, 2, \cdots),$

so that (2.5) becomes

(9.10) $$1 + \binom{3}{2}\frac{1}{2} + \binom{3}{2}\frac{1}{2^2} + \binom{3}{2}\frac{1}{2^3} + \cdots$$

which converges. For this case (8.1) yields

(9.11)
$$p(0) = 1 - \binom{3}{2}\frac{1}{2} + \binom{3}{2}\frac{1}{2^2} - \binom{3}{2}\frac{1}{2^3} + \cdots = \tfrac{1}{2}$$

$$p(1) = \binom{3}{2}\frac{1}{2} - 2!\binom{3}{2}\frac{1}{2^2} + \frac{3!}{2!}\binom{3}{2}\frac{1}{2^3} - \cdots = \tfrac{1}{3}$$

etc.

In this case, the first several coefficients of (4.8) are for $\lambda = 0.75$

(9.12)
$$L_1 = 0, \qquad \frac{L_2}{2!} = .093750, \qquad \frac{L_3}{3!} = .046875, \qquad \frac{L_4}{4!} = .019043$$

$$\frac{L_5}{5!} = .010840, \qquad \frac{L_6}{6!} = .005173, \qquad \frac{L_7}{7!} = .002622, \qquad \cdots .$$

Let us now consider the function (suggested by Prof. C. Wexler)

(9.13) $$p(0) = \frac{5}{3}, \qquad p(r) = (-1)^r \frac{5}{3}\left(\frac{2}{3}\right)^r, \qquad (r = 1, 2, \cdots).$$

For this function

(9.14) $$\sum_{r=0}^{\infty} p(r) = 1, \qquad \sum_{r=0}^{\infty} |\, p(r)\, | = 5$$

(9.15) $$\varphi(z) = \sum_{r=0}^{\infty} z^r \, p(r) = 5/(3 + 2z)$$

(9.16) $$\varphi^{(n)}(1) = \mu_{(n)} = (-1)^n \, n! \, (2/5)^n.$$

In this case (2.5) becomes

(9.17) $$1 + \frac{2}{5} + \left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^3 + \cdots$$

which converges and (8.1) yields

(9.18)
$$p(0) = 1 + \frac{2}{5} + \left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^3 + \cdots = 5/3$$

$$p(1) = -2/5 - 2!(2/5)^2 - \frac{3!}{2!}(2/5)^3 - \cdots = -\frac{5}{3} \cdot \frac{2}{3}$$

etc.

Note that for this case (6.1) or (6.3) are *not* satisfied. Using $\lambda = 1$, it is found that

(9.19) $$L_1 = -1.4, \qquad \frac{L_2}{2!} = 1.06, \qquad \frac{L_3}{3!} = -.5906, \qquad \frac{L_4}{4!} = .2779, \quad \cdots .$$

REFERENCES

[1] A. C. AITKEN, *Statistical Mathematics*, London, 1939, pp. 58–59, pp. 66–67, pp. 76–79.

[2] L. A. AROIAN, "The type B Gram-Charlier series," *Annals of Math. Stat.*, Vol. 8 (1937), p. 183.

[3] C. V. L. CHARLIER, "Über die darstellung willkürlicher Funktionen," *Arkiv. f. Math. Astron. o. Fysik*, Vol. 2 (1905–6), pp. 1–33.

[4] ARNE FISHER, *The Mathematical Theory of Probabilities*, Macmillan, 1923, pp. 271–279.

[5] M. JACOB, "Über die Charlier'sche B-Reihe," *Skand. Aktuarie-tids.*, Vol. 15(1932), pp. 286–291.

[6] CHARLES JORDAN, "Sur la probabilité des épreuves répétées, le théorème de Bernoulli et son inversion," *Bull. Soc. Math. de France*, Vol. 54 (1926), pp. 101–137.

[7] CHARLES JORDAN, *Statistique Mathématique*, Paris, 1927, p. 39.

[8] I. KAPLANSKY, "The asymptotic distribution of runs of consecutive elements," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 200–203.

[9] M. KENDALL, *Advanced Theory of Statistics*, Vol. 1, Griffin, 1943. pp. 154–156.

[10] R. V. MISES, *Wahrscheinlichkeitsrechnung*, Wien. 1931, pp. 265–269.

[11] N. OBRECHKOFF, "Sur la loi de Poisson, la série de Charlier et les équations linéaires aux différences finies du premier ordre a coefficients constants," *Actualités Sci. et Ind.*, Vol. 740 (1938).

[12] H. POLLACZEK-GEIRINGER, "Die Charlier'sche Entwicklung willkürlicher Verteilungen," *Skand. Aktuarie-tids.*, Vol. 11 (1928), pp. 98–111.

[13] H. POLLACZEK-GEIRINGER, "Über die Poissonsche Verteilung und die Entwicklung willkürlicher Verteilungen," *Zeits. f. Angew. Math.*, Vol. 8 (1928), pp. 292–309.

[14] H. L. RIETZ, *Mathematical Statistics*, Chicago, 1927, pp. 60–68, pp. 170–172.

[15] E. SCHMIDT, *Sitzungsberichte der Preuss. Akad. d. Wiss.*, 1928, p. 148.

[16] E. SCHMIDT, "Über die Charlier-Jordansche Entwicklung einer willkürlichen Funktion nach der Poissonschen Funktion und ihren Ableitungen," *Zeits. Angew. Math. Mech.*, Vol. 13 (1933), pp. 139–142.

[17] J. F. STEFFENSEN, "Factorial moments and discontinuous frequency functions," *Skand. Aktuarie-tids.*, Vol. 6 (1923), pp. 73–89.

[18] J. V. USPENSKY, "On Charles Jordan's series for probability," *Annals of Math.*, Vol. 32 (1931), pp. 306–312.

[19] R. P. AGNEW, "Summability of power series," *Amer. Math. Monthly*, Vol. 53 (1946), pp. 251–259.

NOTES

This section is devoted to brief research expository articles on methodology and other short items.

ON SMALL-SAMPLE ESTIMATION

By George W. Brown

Iowa State College

1. Summary. This paper discusses some of the concepts underlying small sample estimation and reexamines, in particular, the current notions on "unbiased" estimation. Alternatives to the usual unbiased property are examined with respect to invariance under simultaneous one-to-one transformation of parameter and estimate; one of these alternatives, closely related to the maximum likelihood method, seems to be new. The property of being unbiased in the likelihood sense is essentially equivalent to the statement that the estimate is a maximum likelihood estimate based on some distribution derived by integration from the original sampling distribution, by virtue of a "hereditary" property of maximum likelihood estimation.

An exposition of maximum likelihood estimation is given in terms of optimum pairwise selection with equal weights, providing a type of rationale for small sample estimation by maximum likelihood.

2. Introduction. In large sample theory of estimation the problems are generally formulated in terms of a random variable $x = (x_1, x_2, \cdots, x_n)$ and a product distribution with, say, a density $g(x|\theta) = f(x_1|\theta)f(x_2|\theta) \cdots f(x_n|\theta)$ where n is permitted to increase without limit. For small sample theory it is sufficient to consider an arbitrary distribution, not necessarily of product form, depending on a parameter θ. For convenience we will assume a distribution density of fixed form $g(x|\theta)$, where x is in Euclidean n-space and θ in Euclidean k-space, $k \leq n$. Granting at the outset that a complete rationale for estimation must be based on considerations like those of Wald [4, 1939] dealing with specified risk functions, it is still a difficult process, in practice, to specify the risk functions and solve the ensuing mathematics problems. It may still be to the point, then, to consider general properties that estimates might be required to have in order to be considered "acceptable", or perhaps even "optimum", over a class of "acceptable" estimates.

In large-sample theory the situation is fairly simple. Consistent estimates have the property that the estimate converges in probability to the true parameter value. "Best" or "optimum" estimates are defined in terms of the order of convergence, or asymptotic variance. All reasonable definitions of "optimum" become asymptotically equivalent, since they all measure essentially the rate of

convergence, so that one might ask for least variance, or least expected absolute deviation, or least expected kth power, without affecting the optimum estimate, in general. Moreover, the consistency property and the optimum properties are in general invariant under simultaneous one-to-one transformation of the parameter and its estimate, i.e., the square of an asymptotically optimum estimate of σ will be an asymptotically optimum estimate of σ^2. Finally, a general estimation method, the method of maximum likelihood, leads to optimum estimates in large samples.

In small samples, on the other hand, the search for corresponding criteria has led to the investigation of best "unbiased" estimates, and the like, where few, if any, of the definitions discussed possess an invariance property under simultaneous one-to-one transformation of the parameter and its estimate.

3. Unbiased estimation. To ensure, in small-sample estimation, that an estimate bears some relation to the parameter it is estimating, it has become the custom to require that an estimate be *unbiased*, which means that the expected value of the estimate agrees with the parameter value. This condition was suggested by the consistency property which is required in large-sample estimation. It ensures, moreover, that the average of a large number of independent estimates made on the same basis will provide a consistent estimate, in the large sample sense. While this consistency property of the average may at times be convenient in practical situations, the fact remains that the problem of estimation from a number of such observations is a different estimation problem, the "best" solution to which need not be the average of the "best" solutions of the original problem corresponding to estimation of θ from a single observation on x, where x has a density $g(x|\theta)$. More to the point, however, is the objection that an unbiased estimate of a parameter does not in general transform into an unbiased estimate when both estimate and parameter are subjected to the same one-to-one transformation. Moreover, one can easily construct situations for which the only acceptable unbiased estimates are clearly inferior from almost any point of view, to estimates which are biased (Girshick, Mosteller and Savage, [1, 1946], and Halmos [2, 1946]).

It may be of interest to consider a few reasonable alternatives to the lack of bias requirement, which seem to accomplish as much as the conventional definition and which, in addition, have an invariance under one-to-one transformation of the parameter and estimate. To avoid confusion, let us attach the qualifying prefix "mean" to the usual unbiased property, so that an estimate will be said to be *mean-unbiased* if its expected value agrees with the parameter value.

Consider as one alternative the following property. An estimate of a one-dimensional parameter θ will be said to be *median-unbiased*, if for fixed θ, the median of the distribution of the estimate is at the value θ, i.e., the estimate underestimates just as often as it overestimates. This requirement seems for most purposes to accomplish as much as the mean-unbiased requirement and has the additional property that it is invariant under one-to-one transformation.

A different alternative requirement which is invariant under transformations is suggested by the definition of unbiased tests of significance (Neyman and Pearson [3, 1936]). Let us say that an estimate is *likelihood-unbiased* if $h(\theta|\theta') \leq h(\theta \mid \theta)$, where the estimate $\hat\theta$ has probability density $h(\hat\theta \| \theta)$. In other words, an estimation method is likelihood-unbiased if estimates in the neighborhood of a given parameter value θ would occur more frequently when the true value is itself θ than when it differs from θ. On intuitive grounds this seems to be an acceptable kind of requirement, applicable to a very general class of estimation problems. It is evident that the assumption of a density plays no important role here; the situation is analogous to the maximum likelihood situation. The property itself is invariant under simultaneous one-to-one transformations of parameter and estimate for the same reason that maximum likelihood estimates are invariant under such transformations, in fact one can readily see that the likelihood-unbiased condition is equivalent to requiring that $\hat\theta$ have such a distribution, as a function of θ, that the maximum likelihood estimate of θ based on $\hat\theta$ will be actually equal to $\hat\theta$. The obvious implication of this fact is that if a function $\phi(x)$ is given (possibly a sufficient statistic for θ) then there is an essentially unique likelihood-unbiased estimate $\hat\theta$ based on ϕ, obtained by finding the maximum likelihood estimate of θ in the distribution of ϕ as a function of θ.

As an example, consider the estimation of σ^2 from a sample of n observations from a normal distribution. Let S^2 be the usual sum of squares, where S^2/σ^2 is distributed like χ^2 on $n-1$ degrees of freedom. Then the only likelihood—unbiased estimate of σ^2 based on S^2 is $S^2/(n-1)$. In this case $S^2/(n-1)$ is also mean-unbiased, a fact which is normally quoted as justification for the division by $n-1$. Curiously enough, it is customary to estimate σ by $\sqrt{S^2/(n-1)}$, even though this is a biased estimate of σ, according to the usual notion of "unbiased", referred to here as "mean-unbiased". On the other hand, $\sqrt{S^2/(n-1)}$ is a perfectly good likelihood-unbiased estimate of σ, by virtue of the invariance under transformations. It might be pointed out, in passing, that the estimate $S^2/(n-1)$ does not have minimum mean square about σ^2, but that the optimum divisor for minimizing the mean square error about σ^2 is $n+1$.

The fact that a likelihood-unbiased estimate is the maximum likelihood estimate based on the distribution of the estimate itself suggest further examination of maximum likelihood estimates. If we define a *simple* estimate as one which completely determines a probability distribution for x, then we have as a theorem, the following:

A simple maximum likelihood estimate $\hat\theta(x)$ is likelihood-unbiased. What this means is essentially that maximum-likelihood is "hereditary", i.e. if $\hat\theta(x)$ maximizes $g(x \mid \theta)$ in a space of n dimensions, and $\hat\theta$ has a derived density $h(\hat\theta \mid \theta)$ in a space of $k \leq n$ dimensions, then $\theta = \hat\theta$ maximizes $h(\hat\theta \mid \theta)$. The proof follows readily from the fact that $h(\hat\theta \mid \theta)$ is obtained by integration of $g(x \mid \theta)$ over all x such that $\hat\theta(x) = \hat\theta$.

The example of estimating σ^2, quoted above, shows that the word "simple" cannot be omitted from the statement above. For example, the simple estimate in the parent distribution is the joint estimate $(x, S^2/n)$ of (m, σ^2) and in fact the joint estimate is likelihood-unbiased. On the other hand, S^2/n is not a simple maximum likelihood estimate, and we observe that S^2/n is not likelihood-unbiased. $S^2/(n - 1)$ is a simple maximum likelihood estimate of σ^2 based on the distribution of S^2 itself, so that $S^2/(n - 1)$ is, as a result, likelihood unbiased.

One can exhibit situations in which the conventional mean-unbiased property is very unnatural, while the likelihood-unbiased property may be quite natural. Consider, for example, the case where σ^2 is to be estimated by use of a χ^2-distributed S^2 with $n - 1$ degrees of freedom, but subject to the condition $\sigma^2 \geq \sigma_0^2$, where σ_0^2 is known in advance. Then the estimate $\sigma^2 = \max [S^2/(n - 1), \sigma_0^2]$ is certainly biased according to conventional definitions, but is nevertheless, likelihood unbiased. To get a mean-unbiased estimate when σ^2 is near to σ_0^2 is impossible except by admitting estimates less than σ_0^2, which is clearly foolish if it is known that $\sigma^2 \geq \sigma_0^2$.

It may be of interest to include a brief discussion of maximum likelihood estimation in terms of pairwise selection of alternatives, providing a sort of optimum property for maximum likelihood estimation in small samples, in addition to the likelihood-unbiased property. Consider a choice to be made between only two alternative values of θ, say θ_0 and θ_1, by dividing the sample space into two regions S_0 and S_1, such that θ_0 is accepted when x falls in S_0 and θ_1 is accepted when x falls in S_1. Then

$$P_{\theta_0}(S_0) + P_{\theta_0}(S_1) = P_{\theta_1}(S_0) + P_{\theta_1}(S_1) = 1.$$

$P_{\theta_1}(S_0)$ is the probability of making the error of accepting θ_0 when $\theta = \theta_1$ and $1 - P_{\theta_0}(S_0)$ is the probability of making the error of accepting θ_1 when $\theta = \theta_0$. If the two errors are weighted equally, it is evident that a "best" test will choose S_0 so as to minimize $P_{\theta_1}(S_0) + 1 - P_{\theta_0}(S_0)$. It is well known that S_0 will minimize the indicated quantity if S_0 consists of all points x such that $g(x \mid \theta_0) \geq g(x \mid \theta_1)$. Thus we may speak of the region S_0 defined by $g(x \mid \theta_0) \geq g(x \mid \theta_1)$ as an *optimum equal risk acceptance region* for θ_0 against θ_1. Now if we transfer our attention to the general estimation problem we see that the maximum likelihood estimate $\hat{\theta}(x)$ is that value of θ which would be accepted by the optimum equal risk acceptance procedure against all other θ's.

REFERENCES

[1] M. A. Girshick, Frederick Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Annals of Math. Stat.*, Vol. 17 (1946), p. 13.

[2] Paul R. Halmos, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), p. 34.

[3] J. Neyman and E. S. Pearson, "Unbiased critical regions of Type A and Type A₁", *Stat. Res. Mem.*, Vol. 1, p. 1.

[4] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses", *Annals of Math. Stat.*, Vol. 10 (1939), p. 299.

A NOTE ON REGRESSION ANALYSIS

By Abraham Wald

Columbia University

1. Introduction. In regression analysis a set of variables $y, x_1, \cdots, x_p$ is considered where y is called the dependent variable and $x_1, \cdots, x_p$ are the independent variables. Let y_α denote the αth observation on y and $x_{i\alpha}$ the αth observation on x_i, $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$. The observations $x_{i\alpha}$ are treated as given constants, while the observations $y_1, \cdots, y_N$ are regarded as chance variables. The following two assumptions are usually made concerning the joint distribution of the variates $y_1, \cdots, y_N$:

(a) The variates $y_1, \cdots, y_N$ are normally and independently distributed with a common unknown variance σ^2.

(b) The expected value of y_α is equal to $\beta_1 x_{1\alpha} + \cdots + \beta_p x_{p\alpha}$ where $\beta_1, \cdots, \beta_p$ are unknown constants.

In some problems it seems reasonable to assume that the regression coefficients $\beta_1, \cdots, \beta_p$ are not constants, but chance variables. This leads to a different probability model for regression analysis and the object of this note is to discuss certain aspects of this model. In what follows in this note we shall make the following assumptions concerning the joint distribution of the chance variables $y_1, \cdots, y_N; \beta_1, \cdots, \beta_p$.

Assumption 1. For given values of $\beta_1, \cdots, \beta_p$ the joint conditional probability density function of $y_1, \cdots, y_N$ is given by

$$(1.1) \qquad \frac{1}{(2\pi)^{N/2} \sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{\alpha=1}^{N} (y_\alpha - \beta_1 x_{1\alpha} - \cdots - \beta_p x_{p\alpha})^2 \right]$$

Assumption 2. The regression coefficients $\beta_1, \cdots, \beta_p$ are independently distributed.

Assumption 3. The regression coefficients $\beta_1, \cdots, \beta_r$, $(r \leq p)$, are normally distributed with zero means and a common variance σ'^2.

The purpose of this note is to derive confidence limits for the ratio $\dfrac{\sigma'^2}{\sigma^2}$. Such confidence limits have been derived by the author [1] for analysis of variance problems assuming that there are only main effects but no interactions. The regression problem treated in the present note is much more general and includes all the analysis of variance problems with or without interactions as special cases.

It should be remarked that Assumptions 2 and 3 do not exclude the case where $\beta_{r+1}, \cdots, \beta_p$ are constants.

2. Derivation of confidence limits for the ratio $\dfrac{\sigma'^2}{\sigma^2}$. Let $b_1, \cdots, b_p$ be the sample estimates of $\beta_1, \cdots, \beta_p$ obtained by the method of least squares. We

shall denote the difference $b_i - \beta_i$ by ϵ_i, $(i = 1, \cdots, p)$. It is known that for given values of $\beta_1, \cdots, \beta_p$ the conditional joint distribution of $\epsilon_1, \cdots, \epsilon_p$ is normal with zero means and variance-covariance matrix $\| c_{ij} \| \sigma^2$ where

$$(2.1) \qquad \| c_{ij} \| = \| a_{ij} \|^{-1}$$

and

$$(2.2) \qquad a_{ij} = \sum_{\alpha=1}^{N} x_{i\alpha} x_{j\alpha}, \qquad (i, j = 1, \cdots, p).$$

Since the conditional distribution of $\epsilon_1, \cdots, \epsilon_p$ does not depend on the values of $\beta_1, \cdots, \beta_p$, the unconditional distribution of $\epsilon_1, \cdots, \epsilon_p$ is the same as the conditional one, and the set of variates $(\beta_1, \cdots, \beta_p)$ is independently distributed of the set $(\epsilon_1, \cdots, \epsilon_p)$. From this and Assumptions 2 and 3 it follows that $b_1, \cdots, b_r$ have a joint normal distribution and that

$$(2.3) \qquad Eb_i = 0, \qquad (i = 1, \cdots, r)$$

and

$$(2.4) \qquad Eb_i b_j = \left(c_{ij} + \delta_{ij} \frac{\sigma'^2}{\sigma^2} \right) \sigma^2, \qquad (i, j = 1, \cdots, r)$$

where $\delta_{ij} = 0$ for $i \neq j$ and $= 1$ for $i = j$.

We shall denote $\dfrac{\sigma'^2}{\sigma^2}$ by λ and the elements of the inverse of $\| c_{ij} + \delta_{ij}\lambda \|$ by $d_{ij}(\lambda)$, i.e.,

$$(2.5) \qquad \| d_{ij}(\lambda) \| = \| c_{ij} + \delta_{ij}\lambda \|^{-1}, \qquad (i, j = 1, \cdots, r).$$

Then the quadratic form

$$(2.6) \qquad Q(\lambda) = \frac{1}{\sigma^2} \sum_{j=1}^{r} \sum_{i=1}^{r} d_{ij}(\lambda) b_i b_j$$

has the χ^2 distribution with r degrees of freedom.

It is known that for any given values of $\beta_1, \cdots, \beta_p, b_1, \cdots, b_p$ the quadratic form

$$(2.7) \qquad Q_a = \frac{1}{\sigma^2} \sum_{\alpha=1}^{N} (y_\alpha - b_1 x_{1\alpha} - \cdots - b_p x_{p\alpha})^2$$

has the χ^2 distribution with $N - p$ degrees of freedom provided that the rank of the matrix $\| x_{i\alpha} \|$ is p. Hence Q_a and $Q(\lambda)$ are independently distributed and the ratio

$$(2.8) \qquad F = \frac{N - p}{r} \frac{Q(\lambda)}{Q_a}$$

has the F-distribution with r and $N - p$ degrees of freedom.

Let F_1 and F_2 be two values chosen so that

$$(2.9) \qquad \text{Prob.} \, \{F_1 \leqq F \leqq F_2\} = c$$

where c is a given positive constant less than 1. Then the set of all values λ for which the inequality

$$(2.10) \qquad F_1 \leq \frac{N - p}{r} \frac{Q(\lambda)}{Q_a} \leq F_2$$

holds forms a confidence set for λ with the confidence coefficient c.

We shall now show that $Q(\lambda)$ is a monotonic function of λ and, therefore, the confidence set determined by (2.10) is an interval. Let $\| g_{ij} \|$, $(i, j = 1, \cdots, r)$, be an orthogonal matrix and let

$$(2.11) \qquad b_i^* = \sum_{j=1}^{r} g_{ij} b_j .$$

It then follows from (2.3) and (2.4) that

$$(2.12) \qquad E(b_i^*) = 0, \qquad\qquad (i = 1, \cdots, r)$$

and

$$(2.13) \qquad E(b_i^* b_j^*) = (c_{ij}^* + \delta_{ij}\lambda)\sigma^2, \qquad\qquad (i, j = 1, \cdots, r)$$

where

$$(2.14) \qquad c_{ij}^* = \sum_{l=1}^{r} \sum_{k=1}^{r} g_{ik} g_{jl} c_{kl} .$$

Let

$$(2.15) \qquad \| d_{ij}^*(\lambda) \| = \| c_{ij}^* + \delta_{ij}\lambda \|^{-1}, \qquad\qquad (i, j = 1, \cdots, r)$$

and put

$$Q^*(\lambda) = \frac{1}{\sigma^2} \Sigma\Sigma \, d_{ij}^*(\lambda) b_i^* b_j^* .$$

It is easy to verify that $Q^*(\lambda)$ is identically equal to $Q(\lambda)$. Hence, to prove the monotonicity of $Q(\lambda)$, it is sufficient to show that $Q^*(\lambda)$ is a monotonic function of λ. Since no restrictions as to the choice of the orthogonal matrix $\| g_{ij} \|$ are made, we shall choose it so that the matrix $\| c_{ij}^* \|$ becomes diagonal, i.e., $c_{ij}^* = 0$ for $i \neq j$, $(i, j = 1, \cdots, r)$. Then

$$(2.16) \qquad d_{ij}^*(\lambda) = 0 \qquad\qquad \text{for } i \neq j$$

and

$$(2.17) \qquad d_{ii}^*(\lambda) = \frac{1}{c_{ii}^* + \lambda} .$$

Hence

$$(2.18) \qquad Q(\lambda) = Q^*(\lambda) = \frac{1}{\sigma^2} \sum_{i=1}^{r} \frac{b_i^{*2}}{c_{ii}^* + \lambda}$$

is a monotonically decreasing function of λ. The confidence set determined by (2.10) is, therefore, an interval.

The upper end point of the confidence interval is the root in λ of the equation

$$(2.19) \qquad \frac{N - p}{r} \frac{Q(\lambda)}{Q_a} = F_1$$

and the lower end point is the root in λ of the equation

$$(2.20) \qquad \frac{N - p}{r} \frac{Q(\lambda)}{Q_a} = F_2.$$

If equation (2.20) has no root, the lower end point of the confidence interval is put equal to zero.

REFERENCE

[1] A. WALD, "On the analysis of variance in case of multiple classifications with unequal class frequencies", *Annals. of Math Stat*, Vol. 12 (1941).

ON THE SHAPE OF THE ANGULAR CASE OF CAUCHY'S DISTRIBUTION CURVES

BY AUREL WINTNER

The Johns Hopkins University

1. Let ξ be a *linear* random variable, that is, a random variable capable of values x represented by points of a line $-\infty < x < \infty$, and suppose, for simplicity, that ξ has a density of probability, $f(x)$. Then, subject to provisos of convergence, the series

$$F(x) = \sum_{n=-\infty}^{\infty} f(x + n)$$

represents a periodic function, of period 1, having the following significance: $F(x)$ is the density of probability of the *angular* random variable, say Ξ, which is obtained if all the states

$$\cdots, \quad \xi - 2, \quad \xi - 1, \quad \xi, \quad \xi + 1, \quad \xi + 2, \cdots$$

of the linear random variable are identified.

In other words, if a circle of unit circumference rolls from $-\infty$ to ∞ on the ξ-line, then every point of the circumference collects the various densities of probability attached to congruent points of the ξ-line, and a state of Ξ represents a point of the circumference. For a detailed study of the mapping $\xi \to \Xi$ or $f \to F$, cf. [2].

According to Poisson's summation formula, the Fourier constants of the periodic function $F(x)$ can be obtained by restricting u in $g(u)$ to an equidistant sequence of discrete values, where $g(u)$ denotes the Fourier transform of $f(x)$; cf., e.g., [5], p. 78 or [9], pp. 477–478.

2. Consider, in particular, the case in which $f(x)$ is the density of a symmetric distribution which is stable in Cauchy's sense. The determination of the totality of these linear densities of probability is due to Lévy [6]. It was shown in [8] that every such $f(x) = f(-x)$ is a decreasing function of $|x|$. As explained in [8], p. 70, this fact makes superfluous one of the axioms occurring in Gauss' postulational approach to "errors of observation."

The purpose of the present note is the deduction of the angular analogue of the fact just quoted. The analogue states that, if $f(x)$ is symmetric and stable, then the corresponding periodic $F(x)$ is decreasing for $0 \leq x \leq \frac{1}{2}$ (and so, for reasons of symmetry, is increasing for $\frac{1}{2} \leq x \leq 1$). This is contained in the italicized statement of §4 below.

In view of Poisson's rule, quoted above, the periodic densities in question can be defined by certain Fourier series representing generalizations of elliptic theta-series. From this point of view, not even the existence (i.e., the *positivity*) of .the periodic densities is obvious, if arbitrary values of the "precision constant" (denoted below by q) are allowed. The difficulties involved are explained in §3.

3. If q and λ are positive constants the first of which is less than 1, then the (even, periodic) function

$$(1) \qquad\qquad \theta_\lambda(x\,;q) = 1 + 2 \sum_{n=1}^{\infty} q^{n^\lambda} \cos nx,$$

where $q^{n^\lambda} > 0$, has derivatives of arbitrarily high order at every real x. It is regular-analytic at every real x if and only if $\lambda > 0$ is replaced by $\lambda \geq 1$, where the sign of equality holds if and only if the analytic continuation (from the x-axis) is not an entire function. In fact, it is known that a Fourier series $\Sigma(a_n \cos nx + b_n \sin nx)$ is that of a function which is regular-analytic at every real x, and has the period 2π, if and only if $|a_n| + |b_n|$ is majorized by a constant multiple of the nth power of a positive constant which is less than 1; and that the latter constant can be chosen arbitrarily small if and only if the analytic continuation does not lead to any singularity (at a $z \neq \infty$).

Since the function (1) tends to 1 uniformly in x as $q \to +0$, if λ is fixed, there belongs to every $\lambda > 0$ a positive $q^* = q^*(\lambda)$ having the property that

$$(2) \qquad\qquad \theta_\lambda(x\,;q) > 0 \text{ for } 0 \leq x < 2\pi$$

if $0 < q < q^*(\lambda)$. It is less obvious that, if q is sufficiently small with reference to λ, say if $0 < q < q^{**}(\lambda)$, then

$$(3) \qquad\qquad \theta_\lambda(x\,;q) \text{ is decreasing for } 0 \leq x \leq \pi \qquad .$$

(hence, increasing for $\pi \leq x < 2\pi$). The existence of such a $q^{**}(\lambda) < \infty$ for every $\lambda > 0$ can be assured as follows:

If $s_n(x)$ denotes the nth partial sum of the Fourier series $\Sigma(\sin nx)/n$, then $s_n(x)$ is positive for $0 < x < \pi$ (Gronwall, Jackson; for a short proof, cf. [4]).

Hence, a partial summation shows that the sum of a sine series, $\Sigma b_n \sin nx$, must be positive for $0 < x < \pi$ if

$$nb_n - (n + 1)b_{n+1} > 0 \text{ and } nb_n \to 0.$$

Since the first derivative of (1) (with respect to x) results by choosing $b_n = -2nq^{n^\lambda}$, it follows that (3) must be true if

$$n^2q^{n^\lambda} - (n + 1)^2 q^{(n+1)^\lambda} > 0$$

holds for $n = 1, 2, \cdots$. But the last inequality is readily seen to be satisfied from $n = 1$ onward if, while λ is fixed, q tends to 0. This proves that $q^{**}(\lambda)$ exists for every $\lambda > 0$.

4. From these deductions alone, it is quite unexpected that (the best values of) both $q^*(\lambda)$ and $q^{**}(\lambda)$ turn out to be independent of λ when

$$(4) \qquad\qquad\qquad 0 < \lambda \leq 2,$$

i.e., that (1) *satisfies both* (2) *and* (3) *for* $0 < q < 1$, *if* (4) *is assumed.* This fact is of statistical significance, since, on the one hand, it is precisely the restriction (4) which is necessary and sufficient for the existence of Cauchy's (symmetric) "stable" distributions (cf. [6], pp. 254–263) and, on the other hand, the reduction (mod 2π) of the densities of these *linear* distributions leads to the functions (1) as angular densities (cf. [9], pp. 477–478); the numerical value of $q(< 1)$ being determined by the "precision" or "dispersion" of the resulting *angular* distributions.

Under the necessary restriction (4), the *linear* analogue of $q^*(\lambda) = 1$ and of $q^{**}(\lambda) = 1$ was proved in [6], pp. 258–263 and in [8], pp. 71–77, respectively. It will remain undecided whether the restriction (4) is necessary in either of the *angular* cases.

5. Suppose that λ has a fixed value in the range (4). Then there exists a monotone function of t, say $\alpha_\lambda(t)$, for which

$$\exp(-u^\lambda) = \int_0^\infty \exp(-u^2 t) \, d\alpha_\lambda(t)$$

is an identity in u, where $0 < u < \infty$ (cf. [1], p. 769, where further references will be found). Hence, a change of variables shows that

$$q^{n^\lambda} = \int_0^\infty q^{tn^2} \, d\alpha_\lambda(t \mid \log q \mid^{1-2/\lambda})$$

is an identity in q and n, where $0 < q < 1$ and $n = 0, 1, 2, \cdots$ (the integration variable is t). Consequently, from (1),

$$\theta_\lambda(x; q) = \int_0^\infty \theta_2(x; q^t) \, d\alpha_\lambda(t \mid \log q \mid^{1-2/\lambda}),$$

where $0 < q < 1$ and $-\infty < x < \infty$. In fact, the legitimacy of the term-by-term integration is obvious from $0 < q < 1$ and $d\alpha_\lambda \geqq 0$ (even though the integrals are improper).

6. Since α_λ is a non-decreasing function, it is clear from the last formula line that both (2) and (3) will be proved for $0 < q < 1$ and for every λ (satisfying (4)), if it is ascertained that both (2) and (3) hold for $0 < q < 1$ when $\lambda = 2$. But the case $\lambda = 2$ of (1) is an elliptic theta-function, for which both properties in question (cf. the diagram in [3], p. 44) are known; a simple proof can be concluded from what, in Hecke's terminology, is the Eulerian factorization of $\theta_2(x \; ; q)$, as follows:

According to Jacobi, the factorization of the case $\lambda = 2$ of (1) is

$$\theta_2(x \; ; q) = \prod_{n=1}^{\infty} (1 - q^{2n})(1 + 2q^{2n-1} \cos x + q^{4n-2})$$

(cf. [7], pp. 64–65). Thus

$$\theta_2(x \; ; q) = c_q \prod_{n=1}^{\infty} P(x + \pi \; ; q^{2n-1}),$$

where

$$c_q = \prod_{n=1}^{\infty} (1 - q^{2n})$$

and

(5) $$P(x \; ; r) = 1 - 2r \cos x + r^2, \qquad (0 < r < 1),$$

hence

$$P(x \; ; r) > 0 \qquad (0 < r < 1).$$

Since $0 < q < 1$, this proves the case $\lambda = 2$ of (2). Furthermore, logarithmic differentiation of the product representation of $\theta_2(x \; ; q)$ gives

$$\theta_2'(x \; ; q) = \theta_2(x \; ; q) \sum_{n=1}^{\infty} P'(x + \pi \; ; q^{2n-1})/P(x + \pi \; ; q^{2n-1}),$$

where $f' = df/dx$; so that, by (5),

$$P'(x + \pi \; ; r) = -2r \sin x.$$

Since $0 < q < 1$, the last three formula lines and the case $\lambda = 2$ of (2) imply that

$$\theta_2'(x \; ; q) < 0 \text{ if } 0 < x < \pi,$$

as claimed by the case $\lambda = 2$ of (3).

This completes the proof of the italicized assertion.

REFERENCES

[1] P. HARTMAN AND A. WINTNER, "On the spherical approach to the normal distribution law," *Amer. Jour. of Math.*, Vol. 62 (1940), pp. 759–779.

[2] E. K. HAVILAND, "On the distribution functions of the reciprocal of a function and of a function reduced mod 1," *Amer. Jour. of Math.*, Vol. 68 (1941), pp. 408–414.

[3] E. JAHNKE AND F. EMDE, *Tables of Functions with Formulae and Curves*, 3d ed. (1938), reprinted by G. E. Stechert and Co., New York.

[4] E. LANDAU, "Ueber eine trigonometrische Ungleichung," *Math. Zeit.*, Vol. 37 (1933), p. 36.

[5] H. LEBESGUE, *Leçons sur les séries trigonométriques*, Paris, 1905.

[6] P. LÉVY, *Calcul des probabilités*, Paris, 1925.

[7] H. WEBER, *Elliptische Funktionen und algebraische Zahlen*, Braunschweig, 1891.

[8] A. WINTNER, "On a class of Fourier transforms," *Amer. Jour. of Math.*, Vol. 58 (1936), pp. 45–90 and p. 425.

[9] F. ZERNIKE, "Wahrscheinlichkeitsrechnung und mathematische Statistik," *Handbuch der Physik*, Vol. 3 (1928), pp. 419–492.

A NOTE ON THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS

BY G. E. ALBERT

U. S. Naval Ordnance Plant, Indianapolis

1. Introduction. Let $\{z_i\}$, $(i = 1, 2, 3, \cdots)$, be a sequence of real valued random variables identically distributed according to the cumulative distribution function $F(z)$. Define the sums $Z_N = z_1 + z_2 + \cdots + z_N$ for every positive integer N. Choose two positive constants a and b and define the random variable n as the smallest integer N for which one of the inequalities $Z_N \geqq a$ or $Z_N \leqq -b$ holds. The notations $P(u \mid F)$ and $E(u \mid F)$ will denote the probability of u and its expectation respectively assuming that F is the distribution of the z_i.

Wald [1] has established the results contained in the following lemmas.

LEMMA 1. *If the variance of $F(z)$ is positive, $P(n < \infty \mid F)$ equals one.*

LEMMA 2. *If there exists a positive number δ such that $P(e^z < 1 - \delta \mid F) > 0$ and $P(e^z > 1 + \delta \mid F) > 0$ and if the moment generating function $\varphi(t) = E(e^{zt} \mid F)$ exists for all real values of t, then $\varphi(t)$ has one and only one minimum at some finite value $t = t_0$. Moreover, $\varphi''(t) > 0$ for all real values of t.*

It is the purpose of this note to establish the following extension of the validity of certain results given by Wald [1], [2].

THEOREM.[1] *Under the conditions of Lemma 2 the identity*

$$(1) \qquad E\{e^{z_n t}[\varphi(t)]^{-n} \mid F\} = 1$$

---

[1] Wald's results show (1) to be valid for all complex t in the domain over which $\mid \varphi(t) \mid \geqq 1$ and the validity of the differentiation clause for all real t in that domain. The importance of the present extension arises from the fact that, if $E(x \mid F) \neq 0$, then $0 < \varphi(t) < 1$ on a certain interval of the real axis.

is valid and may be differentiated with respect to t under the expectation sign any number of times for all real values of t.

PROOF. The notation t_0 will be used consistently to denote the t value at which $\varphi(t)$ has its minimum.

The proof of the theorem follows Wald's methods quite closely and certain of the results given in [1] and [2] will be used here without discussion.

Consider first the validity of (1). For an arbitrary positive integer N let P_N be the probability $P(n \leq N \mid F)$ and let $E_N(u \mid F)$ and $E_N^*(u \mid F)$ denote the conditional expectations of u subject to the respective conditions $n \leq N$ and $n > N$. Wald [1] has shown that for any finite real value of t

$$(2) \qquad P_N E_N\{e^{Z_n t}[\varphi(t)]^{-n} \mid F\} + (1 - P_N)[\varphi(t)]^{-N} E_N^*\{e^{Z_N t} \mid F\} = 1.$$

Since $\lim_{N \to \infty} P_N E_N\{[\varphi(t)]^{-n} \exp(Z_n t)\}$ is the left member of the identity (1), it suffices to demonstrate that

$$(3) \qquad \lim_{N \to \infty} (1 - P_N)[\varphi(t)]^{-N} E_N^*\{e^{Z_N t} \mid F\} = 0$$

for all real values of t.

Since $1 - P_N$ tends to zero with increasing N and the expected value E_N^* involved in (3) is bounded independently of N for any fixed t, the only source of difficulty in proving (3) lies in the fact that $\varphi(t)$ may be less than unity on an interval of the real axis. That difficulty is easily avoided by the following device. Define the function

$$(4) \qquad G(x) = [\varphi(t_0)]^{-1} \int_{-\infty}^{x} e^{z t_0} \, dF(z).$$

Obviously $G(x)$ is a distribution function whose moment generating function $\psi(t)$ exists for all real t. Its mean is zero and its variance is positive as will be seen from the equations $E(x \mid G) = \varphi'(t_0)/\varphi(t_0)$ and $E(x^2 \mid G) = \varphi''(t_0)/\varphi(t_0)$. It follows that $\psi(t)$ is never less than unity for real values of t.

Let Ω denote the space of all $z_1, \cdots, z_N$ and let $\Omega(n > N)$ be that subset of Ω on which $n > N$. One has

$$(1 - P_N)[\varphi(t)]^{-N} E_N^*\{e^{Z_N t} \mid F\}$$

$$= \frac{\int_{\Omega(n>N)} e^{Z_N t} \, dF(z_1) \cdots dF(z_N)}{\int_{\Omega} e^{Z_N t} \, dF(z_1) \cdots dF(z_N)} = \frac{\int_{\Omega(n>N)} e^{Z_N(t-t_0)} \, dG(z_1) \cdots dG(z_N)}{\int_{\Omega} e^{Z_N(t-t_0)} \, dG(z_1) \cdots dG(z_N)}$$

$$= (1 - Q_N)[\psi(s)]^{-N} E_N^*\{e^{Z_N s} \mid G\}$$

where $s = t - t_0$ and $Q_N = P(n \leq N \mid G)$. By Lemma 1, $1 - Q_N$ tends to zero as N is increased. Thus, since $\psi(s) \geq 1$ for all real t and the expected value $E_N^*\{e^{Z_N s} \mid G\}$ is bounded independently of N for a fixed t, the equation (3) holds for all real t.

The differentiability clause of the theorem requires the following modification of a very powerful theorem due to Charles Stein [3].

LEMMA 3. *Under the conditions of Lemma 2, if the minimum $\varphi(t_0)$ of $\varphi(t)$ is less than unity, there exists a positive number t_1 such that*

$$(5) \qquad E\{\exp [nt_1 - n \log \varphi(t_0)] \mid F\} < \infty.$$

PROOF. If G is the distribution of the z_i, by Stein's theorem there exists a positive number t_1 such that $E(e^{nt_1} \mid G)$ is finite. Let $\Omega(n = N)$ denote the subset of Ω on which $n = N$. Then

$$P(n = N \mid G) = \int_{\Omega(n=N)} dG(z_1) \cdots dG(z_N)$$

$$= [\varphi(t_0)]^{-N} \int_{\Omega(n=N)} e^{t_0 Z_N} dF(z_1) \cdots dF(z_N)$$

$$\geqq P(n = N \mid F) \exp [\min\{at_0, -bt_0\} - N \log \varphi(t_0)].$$

It follows that

$$E\{\exp [nt_1 - n \log \varphi(t_0)] \mid F\} \leqq E\{e^{nt_1} \mid G\} \exp [-\min\{at_0, -bt_0\}]$$

and the lemma is proved.

To continue with the theorem, Wald's proof [2] suffices for the case in which $\varphi(t_0) \geqq 1$. Attention will be given only to the case $\varphi(t_0) < 1$. As pointed out in section 2 of [2], the differentiability clause of the theorem will be established if it can be shown that for any finite interval I of the real axis and any pair of integers r_1 and r_2 there exists a function $D_{r_1 r_2}(Z_n, n)$ such that for all t in I one has

$$(6) \qquad D_{r_1 r_2}(Z_n, n) \geqq \mid n^{r_1} Z_n^{r_2} e^{Z_n t}[\varphi(t)]^{-n} \mid$$

and

$$(7) \qquad E\{D_{r_1 r_2}(Z_n, n) \mid F\} < \infty.$$

On referring to Wald's proof and using the inequality $-\log \varphi(t) \leqq -\log \varphi(t_0)$ for all t in I, it is seen that there exists a constant C and a positive number t_2 such that the function

$$D_{r_1 r_2}(Z_n, n) \equiv Cn^{r_1}[\varphi(t_0)]^{-n}(e^{Z_n t_2} + e^{-Z_n t_2})$$

satisfies (6) for all t in I. To establish (7) use the inequalities (2.4) and (2.6) in Wald [2] to obtain

$$E\{D_{r_1 r_2}(Z_n, n) \mid F\}$$

$$(8) \qquad = C \sum_{N=1}^{\infty} P(n = N \mid F) N^{r_1}[\varphi(t_0)]^{-N} E_{n=N}\{e^{Z_n t_2} + e^{-Z_n t_2} \mid F\}$$

$$\leqq C\{e^{at_2} l(t_2) + e^{-bt_2} l(-t_2)\} E\{\exp [r_1 \log n - n \log \varphi(t_0)] \mid F\}.$$

That (7) is indeed satisfied now follows from (5) and the finiteness of the function $l(t)$ since for a large enough integer M one has

$$\sum_{N=M}^{\infty} P(n = N \mid F) \exp [r_1 \log N - N \log \varphi(t_0)]$$

$$\leqq \sum_{N=M}^{\infty} P(n = N \mid F) \exp [Nt_1 - N \log \varphi(t_0)] < \infty.$$

Thus the expected value on the extreme right in (8) is finite. This completes the proof of the theorem.

REFERENCES

[1] A. WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–285.
[2] A. WALD, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 493–496.
[3] CHARLES STEIN, "A note on cumulative sums," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.

A SIGNIFICANCE TEST AND ESTIMATION IN THE CASE OF EXPONENTIAL REGRESSION

By D. S. VILLARS[1]

United States Rubber Company, Passaic, N. J.

1. Introduction. The principal problem under consideration in this note may be described as follows. Consider a variate, z, whose distribution for a given value of a fixed variate, t, is:

$$(1.1) \qquad f(z \mid t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(z-a+be^{-kt})^2/2\sigma^2}$$

where a, b, and k are real-valued parameters. The regression of z on t is exponential, for it follows from (1.1) that the expected value of z, given t, is:

$$(1.2) \qquad E(z \mid t) = a - be^{-kt}.$$

On the basis of a random sample $0_N(z_1, t_1 ; z_2, t_2 ; \cdots ; z_N, t_N)$ it is desired to test whether $k = 0$ or ∞. The problem of "fitting" a curve, $z = a - be^{-kt}$, to the sample (i. e. of estimating a, b, and k from the sample) will also be treated.

As an illustration of how the statistical problems described above arise in

practice, let us consider a typical situation in industrial chemistry. Let the quantity, z, be a property of a latex and let the quantity, t, be time. Suppose, furthermore, that measurement of t is without error but that measurement of z is subject to error; let it be assumed that the observed value in a measurement of z is a variate having a normal (Gaussian) distribution about the "true value," $E(z)$. On basis of N independent measurements, $z_1, z_2, \cdots, z_N$ of z at times, $t_1, t_2, \cdots, t_N$, respectively, the experimenter may wish to test the hypothesis that $k = 0$ or ∞. If this hypothesis is true the suspected exponential relation between z and t does not hold; in this case $E(z)$ is a constant $(a - b,$ or $a)$ and estimation of the constant from the data is quite straightforward. If the data conflict with the hypothesis that $k = 0$ or ∞, the experimenter may wish to estimate the parameters, a, b, and k (i. e., "fit" the curve, $z = a - be^{-kt}$, to the data).

The problems considered in this note will be treated only for the case where N is an even integer (≥ 6) and the times $t_1, t_2, \cdots, t_N$ at which measurements of z are made are such that

$$(1.3) \qquad t_{2\alpha} - t_{2\alpha-1} = \Delta, \qquad \text{a constant, } (\alpha = 1, 2, \cdots, n = N/2).$$

The odd time intervals, $t_3 - t_2, t_5 - t_4$, etc. do not have to be equal.

2. Test of the hypothesis that $k = 0$ or ∞. The space, say Ω, of admissible values of the parameters in (1.1) is: $\sigma^2 > 0, -\infty < a, b, k < +\infty$. Under the null hypothesis the admissible values of the parameters lie in a subspace of Ω, say ω, specified as follows: $\sigma^2 > 0, -\infty < a, b < +\infty, k = 0,$ or ∞.

Let $y_j = z_{2\alpha}$ and $x_j = z_{2\alpha-1}$, $(\alpha = 1, \cdots, n = N/2)$. From (1.1) and (1.3) it follows that the n pairs x_j, y_j are normally and independently distributed with common variance, σ^2, that x_j and y_j are independent $(j = 1, 2, \cdots, n)$, and that

$$(2.1) \qquad \qquad \nu_j = h + m\mu_j$$

where $\nu_j = E(y_j)$, $\mu_j = E(x_j)$, $h = a(1 - e^{-k\Delta})$, and $m = e^{-k\Delta}$. The space, Ω', of admissible values of the parameters in the joint distribution of x_j, y_j, $(j = 1, \cdots, n)$, is: $\sigma^2 > 0, \nu_j = h + m\mu_j, -\infty < h < +\infty, -\infty < \mu_j, \nu_j < +\infty; 0 \leq m < \infty$. The subspace of Ω', say ω', associated with the null hypothesis is: $\sigma^2 > 0, \nu_j = \mu_j = c$, where $c = a - b$ or a according as $k = 0$ or ∞. In Ω', the expected values of x and y lie on a line; in ω' they lie in a single point. It is clear that by transforming the original sample $0_N(z_1, t_1, \cdots, z_N, t_N)$ to a sample $0_n(x_1, y_1; \cdots; x_n, y_n)$ we have reduced the original problem to the familiar problem of linear regression in which there is "error in both variates".

The slope of the "line of best fit" to the sample points $(x_1, y_1; \cdots; x_n, y_n)$ is [1]:

$$(2.2) \qquad \hat{m} = [S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}]/2S_{xy}$$

where

$$S_{xx} \equiv \sum_{1}^{n} (x_j - \bar{x})^2$$

$$S_{xy} \equiv \sum_{1}^{r} (x_j - \bar{x})(y_j - \bar{y})$$

$$S_{yy} \equiv \sum_{1}^{n} (y_j - \bar{y})^2$$

$$\bar{x} \equiv \sum_{1}^{n} x_j/n$$

$$\bar{y} \equiv \sum_{1}^{n} y_j/n$$

($\hat{m}$ is an estimate of m in (2.1)). Since $m = e^{-k\Delta}$ (where k and Δ are real), it is intuitively clear that when m is non-positive the sample 0_Δ does not conflict with the null hypothesis. The null hypothesis can be tested by means of the statistic [2, 144]

(2.3) $$F' = \frac{S_{xx} + 2mS_{xy} + m^2 S_{yy}}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}.$$

The null hypothesis is rejected if $\hat{m}$ is positive and F' is large. Percentage points of the distribution of F' are given in [2, 146] for $n = 3$ (1) 15 (5) 30, 40, 60, 120 and for significance levels, 0.001, .01, .05, .10, and .20. These significance levels, however, were computed for use in cases where the sign of $\hat{m}$ was irrelevant. It happens that to test the null hypothesis under consideration in this problem at a significance level α we should use a critical value of F' (given in [2]) corresponding to a significance level 2α. The reason for this is that when the null hypothesis is true the quantities m and F' are independent and the probability that $\hat{m}$ is positive is $\frac{1}{2}$—thus the chance of rejecting the null hypothesis is $\frac{1}{2}(2\alpha) = \alpha$.

3. Estimation of a, b, and k. If the data do not support the hypothesis that $k = 0$ or ∞, the experimenter may wish to estimate a, b, and k. General alternative methods of estimating these parameters will now be considered.

(1) Estimate a, b, and k from 0_N by the method of least squares; *i.e.*, solve the simultaneous equations $\partial S/\partial a = 0$, $\partial S/\partial b = 0$, and $\partial S/\partial k = 0$ for a, b, and k, where

(3.1) $$S = \sum_{i=1}^{N} (z_i - a + be^{-kt_i})^2.$$

The value of k obtained by this method of estimation will not in general be the same as that computable from $\hat{m}$ in (2.2) and used for the significance testing.

(2) Estimate k by means of (2.2) and the relation $m = e^{-k\Delta}$, then substitute this estimate into S of (3.1) and estimate a and b by means of least squares.

(3) Estimate k as in (2) and choose, as an estimate of a, the intercept of the "line of best fit" for 0_n. Then substitute these estimates of a and k into (3.1) and estimate b by means of least squares. In this case the estimate of b comes out to be:

$$(3.2) \qquad \hat{b} = \sum_1^N e^{-\hat{k}t_i}(\hat{a} - z_i) \bigg/ \sum_1^N e^{-2\hat{k}t_i}$$

where $\hat{a}$ and $\hat{k}$ are the estimates of a and k.

If the values, $t_1, t_2, \cdots, t_N$ are such that $t_{i+1} - t_i = \Delta, (i = 1, 2, \cdots, N - 1)$, the following estimation procedure might be used.

(4) Let

$$y_i = z_{i+1}$$
$$(i = 1, 2, \cdots, N - 1),$$
$$x_i = z_i$$

and treat the $(N - 1)$ pairs of values $(x_1, y_1; \cdots; x_{N-1}, y_{N-1})$ as a sample of size $(N - 1)$. Using this sample, estimate k, a, and b in a manner similar to that in (2) or (3). It should be noted that this sample is not a random sample owing to the dependence among the $(N - 1)$ elements.

The procedure in alternative (1) is very laborious and time-consuming. The procedure in (2) and (3) can be carried out quickly and easily. In (1) the method of least squares yields the same results as would be obtained from application of the method of maximum likelihood. Examples of estimation by procedures (3) and (4) are given in the next section.

4. Example.

The accompanying table lists experimentally observed values of a property of a latex obtained at biweekly intervals. Using the first, third, etc., quantities as x_i and the remaining ones as y_i, the sums of squares and products of deviations are found to be:

$$S_{xx} = .035510 \qquad\qquad \bar{x} = 0.9195$$

$$S_{xy} = .025645$$

$$S_{yy} = .023414 \qquad\qquad \bar{y} = .9365.$$

Substituting these values in equation (2.2) and computing the other constants from equation (2.1) we get: $m = 0.791596$, $a = 1.0009$, and $k = 0.1168$. The F' ratio is (2.3) 17.03. Entering Table I of [2], we find that for eight point pairs a value of $F' = 16.5$ may be expected only one time in one hundred. On excluding the possibility of negative values of m, this corresponds to the 0.5% significance level. The exponential relationship is thus concluded to be highly significant.

Evaluation of b by equation (3.2), method 3, gives 0.2560, if all 16 values are used. The equation calculated from the data is thus:

$$(4.1) \qquad z = 1.0009 - 0.2560\, e^{-0.1168t}.$$

The alternative procedure, method 4, would be to use all the z_i points for the estimation of a and k. This leads to the following values of the computation quantities:

$$S_{xx} = \sum_{i=1}^{16} x_i^2 - x_{16}^2 = 0.052374; \qquad \bar{x} = 0.9223$$

$$S_{xy} = \sum_{i=1}^{15} x_i x_{i+1} = .036924$$

$$S_{yy} = \sum_{i=1}^{16} x_i^2 - x_1^2 = .035436; \qquad \bar{y} = .9381.$$

Note that the difference $S_{yy} - S_{xx}$ used in the formula for m cancels out all intervening squares between the first and last.

$$S_{yy} - S_{xx} = x_1^2 - x_{16}^2.$$

TABLE I

t weeks	z_i	t weeks	z_i	t weeks	z_i	t weeks	z_i
1	.776	9	.939	17	.942	25	.955
3	.852	11	.904	19	.938	27	.993
5	.850	13	.930	21	.979	29	.985
7	.869	15	.948	23	.975	31	1.013

However, the data excluded thereby are in effect included in the new S_{xy}.

The final values obtained by the fourth procedure are: $m = 0.796596$, $a = 1.0000$, and $k = 0.1137$. The writer does not know whether the peculiar transference of data from $S_{yy} - S_{xx}$ to S_{xy} characteristic of procedure 4 improves the accuracy of the fit or hurts it. It is his personal preference to use procedure 3.

5. Acknowledgement. The writer wishes to acknowledge with thanks his gratitude to Drs. T. W. Anderson, Jr. and David F. Votaw, Jr. for many suggestions and discussions concerning this problem and for much help in clarifying the presentation of the concepts.

REFERENCES

[1] CHARLES H. KUMMELL, *The Analyst* (Des Moines), Vol. 6 (1879), pp. 97–105.
[2] D. S. VILLARS AND T. W. ANDERSON, JR., "Some significance tests for normal bivariate distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 141–148.

ON THE POWER EFFICIENCY OF A t-TEST FORMED BY PAIRING SAMPLE VALUES

By John E. Walsh

Princeton University

1. Introduction. Consider two equal sized samples, one from a normal population with mean μ and the other from a normal population with mean ν. Let $x_1, \cdots, x_n$ be the sample values from the population with mean μ and $y_1, \cdots, y_n$ the values from the population with mean ν. If the two populations have the same variance and the two samples are independent, the most powerful tests for comparing μ and ν using these samples (one-sided and symmetrical two-sided) are based on the statistic

$$t_2 = \frac{[\bar{x} - \bar{y} - (\mu - \nu)]\sqrt{n(n-1)}}{\sqrt{\sum_{1}^{n}(x_i - \bar{x})^2 + \sum_{1}^{n}(y_i - \bar{y})^2}},$$

which has a Student t-distribution with $2n - 2$ degrees of freedom. Tests based on t_2 also have the desirable property of being invariant under permutation of the data in each sample.

Sometimes, however, it is useful to combine the sample values in the form

$$z_i = (x_i - y_i), \qquad (i = 1, \cdots, n).$$

Examples:

(a). When the samples are independent but it is not known that the two populations have the same variance (Behrens-Fisher problem).

(b). When there may be correlation between x_i and y_i, $(i = 1, \cdots, n)$, this correlation being the same for each value of i (i.e. x_i is independent of y_j if $i \neq j$ while each pair x_i, y_i, $(i = 1, \cdots, n)$, has the same normal bivariate distribution).

In both (a) and (b) the z_i are independently normally distributed with the same variance and mean $\mu - \nu$.

The Student t-test for comparing μ and ν using the z_i is based on the statistic

$$t_1 = \frac{[\bar{z} - (\mu - \nu)]\sqrt{n(n-1)}}{\sqrt{\sum_{1}^{n}(z_i - \bar{z})^2}} = \frac{[\bar{x} - \bar{y} - (\mu - \nu)]\sqrt{n(n-1)}}{\sqrt{\sum_{1}^{n}[x_i - y_i - (\bar{x} - \bar{y})]^2}},$$

which has a Student t-distribution with $n - 1$ degrees of freedom. These tests are not invariant under permutation of the data in each sample.

If it is true that all the sample values are independently distributed with the same variance σ^2, efficiency will be lost by using the test based on t_1 instead of the most powerful test based on t_2. The purpose of this note is to determine the power efficiency of the tests based on t_1 as compared with the corresponding tests based on t_2 for this case.

TABLE I

Power Function Values for the t_1 and t_2 Tests

Test		Approx. Efficiency		Approx. Values of Power Function			
				$\delta = \frac{1}{2}$	$\delta = 1$	$\delta = 1\frac{1}{2}$	$\delta = 2$
t_1	6	87%	.05	.276	.674	.933	.994
t_2	5.2		.05	.275	.672	.932	.994
t_1	6	82.5%	.025	.159	.486	.822	.970
t_2	4.95		.025	.160	.488	.823	.970
t_1	8	90%	.05	.355	.812	.985	
t_2	7.2		.05	.354	.813	.985	
t_1	8	86.5%	.025	.226	.674	.952	.998
t_2	6.9		.025	.225	.675	.951	.998
t_1	8	82%	.01	.112	.458	.843	.983
t_2	6.55		.01	.112	.457	.842	.983
t_1	10	92%	.05	.425	.898	.997	
t_2	9.2		.05	.425	.897	.997	
t_1	10	90%	.025	.289	.802	.988	
t_2	9		.025	.290	.803	.988	
t_1	10	85.5%	.01	.159	.626	.950	.999
t_2	8.55		.01	.159	.627	.950	.999
t_1	15	95.5%	.05	.579	.980		
t_2	14.3		.05	.579	.980		
t_1	15	93%	.025	.437	.950	1.000	
t_2	13.95		.025	.437	.949	1.000	
t_1	15	90%	.01	.278	.876	.998	
t_2	13.5		.01	.278	.876	.998	
t_1	25	98%	.05	.784	.999		
t_2	24.5		.05	.784	.999		
t_1	25	96%	.025	.670	.998		
t_2	24		.025	.670	.998		
t_1	25	94.5%	.01	.514	.992		
t_2	23.7		.01	.514	.992		

Consideration is limited to one-sided tests, which is not a serious limitation since any two-sided test can be considered as a combination of two one-sided tests. Table II contains approximate power efficiencies of one-sided tests for $n \geq 4$ at the significance levels $\alpha = .05, .025, .01$.

It is found that the efficiency of the t_1 test increases with the sample size but is high even for small size samples.

2. Outline of computations. The method of obtaining power efficiencies used here will be that outlined in [1]. Essentially this consists in computing the power function for the test based on t_1 and then adjusting the sample size for the corresponding test based on t_2 until its power function is approximately the same as for the t_1 test. The ratio of the sample size (perhaps fractional) of the adjusted t_2 test to that of the t_1 test is called the power efficiency of the t_1 test. Intuitively this efficiency measures the fraction of the total available information which is being used when the t_1 test is applied (since the t_2 test is most powerful).

TABLE II

Approximate Power Efficiencies for Given n and α

α \ n	4	5	6	7	8	9	10	15	25	∞
.05	82.5%	85%	87%	88.5%	90%	91%	92%	95.5%	98%	100%
.025	77%*	80%*	82.5%	84.5%	86.5%	88.5%	90%	93%	96%	100%
.01	73%	75.5%	78%	80%	82%	84%	85.5%	90%	94.5%	100%

* These values were obtained by comparison with the corresponding values for $\alpha = .05$ and .01.

It is easily seen from symmetry that a one-sided t_1 test of $\mu < \nu$ has the same power efficiency as the corresponding one-sided t_1 test of $\mu > \nu$. Thus it is sufficient to consider the one-sided tests of $\mu > \nu$.

The power function is found as a function of the parameter δ, where

$$\delta = \frac{\mu - \nu}{\sigma \sqrt{2}}.$$

Most of the approximate power efficiencies were determined by using the normal approximation given in [2] to compute the power function values. This approximation was used for fractional values of n. Table I contains the results of these computations for one-sided tests of $\mu > \nu$.

Exact values of the power function for integral values of n and $\alpha = .05, .01$ can be found from the tables in [3]. A comparison of the power function values obtained from the normal approximation with these exact values shows that, for $n \leq 6, \alpha = .01$ and $n \leq 4, \alpha = .05, .025$, the approximation underestimates the true values for small δ and overestimates for large δ. Although this combination of underestimation and overestimation tends to cancel out in the determina-

tion of power efficiencies, so that little error in power efficiencies would be expected if the approximation were used for $n = 6$, $\alpha = .01$ or $n = 4$, $\alpha = .05$, the efficiencies given in Table II for $n = 4$, $\alpha = .05$ and $n = 4, 6$, $\alpha = .01$ were obtained from the exact values by graphical interpolation and cross-interpolation.

Power efficiencies were not considered for $n < 4$ because of the difficulties of interpolation and the inexactness of the normal approximation in this range.

For $n = \infty$, t_1 and t_2 both have a normal distribution with zero mean and unit variance. Thus the power efficiency is 100% at all significance levels for this case.

These computations furnish approximate power efficiencies for $n = 6, 8, 10, 15, 25, \infty$ at $\alpha = .05, .025, .01$, and for $n = 4$ at $\alpha = .05$ and $.01$. The other approximate power efficiencies listed in Table II were obtained by graphical interpolation from these values.

The results of this note can be roughly summarized for $n \leq 15$ by stating that of the $2n$ sample values

    (i). approximately 1.6 values are lost at the 5% significance level,

    (ii). approximately 2.1 values are lost at the 2.5% significance level,

    (iii). approximately 2.8 values are lost at the 1% significance level, if the tests based on t_1 are used instead of the corresponding tests based on t_2. Examination of Table I shows that the number of sample values lost decreases as n increases for $n > 15$.

REFERENCES

[1] JOHN E. WALSH, "On the power function of the sign test for slippage of means", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 360, 361.

[2] N. L. JOHNSON AND B. L. WELCH, "Applications of the non-central t-distribution", *Biometrika*, Vol. 31 (1940), p. 376.

[3] J. NEYMAN, "Statistical problems in agricultural experimentation", *Roy. Stat. Soc. Suppl.*, Vol. 2 (1935), pp. 131, 132.

NOTE ON THE LIAPOUNOFF INEQUALITY FOR ABSOLUTE MOMENTS

By MAURICE H. BELZ

The University of Melbourne

For a variate x measured from the mean of the population, the absolute moment of order r is defined by

$$\nu_r = \int_{-\infty}^{\infty} |x|^r \, dF(x),$$

where $F(x)$ is the cumulative distribution function. Treating r as continuous, we have

$$\frac{d\nu_r}{dr} = \int_{-\infty}^{\infty} |x|^r \log_e |x| \, dF(x),$$

the integral on the right existing if ν_{r+1} exists.

Write $y = \log_e \nu_r$. Then we have

$$\nu_r \frac{dy}{dr} = \int_{-\infty}^{\infty} |x|^r \log_e |x| \, dF(x),$$

$$\nu_r^2 \frac{d^2 y}{dr^2} = \int_{-\infty}^{\infty} |x|^r \, dF(x) \cdot \int_{-\infty}^{\infty} |x|^r \log_e^2 |x| \, dF(x) - \left\{ \int_{-\infty}^{\infty} |x|^r \log_e |x| \, dF(x) \right\}^2$$

$\geqq 0$, by Schwarz's inequality.
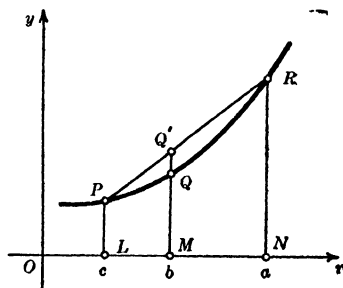
Fig. 1

It follows that the function y is convex (or exceptionally a straight line), and, on referring to the figure, it appears that

(1) $$MQ \leqq MQ'$$

for all chords PR. If the abscissae of the points L, M, N are c, b, a, respectively, where $c \leqq b \leqq a$, the inequality (1) leads at once to the relation

$$\log_e \nu_b \leqq \frac{a - b}{a - c} \log_e \nu_c + \frac{b - c}{a - c} \log_e \nu_a.$$

Hence

$$\nu_b^{a-c} \leqq \nu_c^{a-b} \nu_a^{b-c},$$

which is the usual form of the Liapounoff Inequality.

REMARK ON THE NOTE "A GENERALIZATION OF WARING'S FORMULA"

By T. N. E. Greville

U. S. Public Health Service

Before submitting for publication the note "A generalization of Waring's formula," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 218–219 the author made a diligent effort to ascertain, through correspondence with mathematicians and

actuaries both in this country and abroad, whether the generalized formula in question had been previously published, and none of the authorities communicated with knew of its prior publication. However, it has now come to his attention that the formula was published in essentially the same form by Hermite in the article "Sur la formule d'interpolation de Lagrange", *Journal für die Reine und Angewandte Mathematik* ("Crelle's Journal"), Vol. 84 (1878), pp. 70–79.

ABSTRACTS OF PAPERS

1. Estimation of Parameters in Truncated Pearson Frequency Distributions. A. C. COHEN, University of Georgia.

Given a truncated univariate Pearson frequency distribution, parameters of the complete distribution are required. Karl Pearson and Alice Lee, (*Biometrika*, Vol. 6 (1915), pp. 59–69) and R. A. Fisher, (*Introduction to Mathematical Tables*, Vol. 1, British Assn. Adv. Sci., 1931, pp. xxvi–xxxv), obtained solutions of the truncated normal distribution with a single *tail* missing. The present paper presents three general methods of solution applicable to any of the Pearson distributions. The first utilizes moments of a higher order than are required to characterize corresponding complete distributions. The order of the highest moment required is increased by one for each missing *tail*. The second method, applicable when only a single *tail* is missing, utilizes the terminal ordinate at the point of truncation and moments of the same order as required to characterize the complete distribution. The terminal ordinate is evaluated by successive approximations. The third method utilizes only the first two moments, but requires that the given distribution be further truncated and that moments be computed both before and after the additional truncations. This latter method can also be applied to complete distributions to avoid direct computation of third and fourth order moments.

2. Distribution of a Root of Determinantal Equation. D. N. NANDA, University of North Carolina.

The joint distribution of the roots of a determinantal equation was given by P. L. Hsu in 1939 and the distribution of any one of the roots was studied by S. N. Roy. The present paper, however, gives a different method of working out the distribution of any root, specified by its place in a monotonic arrangement. This method enables us to express the distribution of a root of a certain determinantal equation in terms of a linear combination of products of incomplete beta integrals and in terms of the distribution of a root of lower-order determinantal equations.

3. The Power of Certain Non-Parametric Tests of Independence. WASSILY HOEFFDING, University of North Carolina.

Several tests of independence have been proposed which are based on statistics depending only on the ranks of the sample values. Under the hypothesis H_0 of independence the distribution of such statistics does not depend on the form of the parent distribution. Two of these statistics, Spearman's rank correlation coefficient and Lindeberg-Kendall's statistic based on the number of inversions in the permutation of the ranks, are shown to be asymptotically normally distributed in samples from any population (the limiting normal distribution being singular in certain degenerate cases). The asymptotic distribution of these coefficients reveals that the corresponding tests of independence are inconsistent (in the sense that the probability of rejecting H_0 does not necessarily tend to 1 if H_0 is not true), and at least one of them is biased in the limit. It can be shown that at least for some sample sizes and some sizes of the critical region there do not exist unbiased tests of independence based on ranks. But there do exist rank tests of independence which are consistent, and hence unbiased in the limit. Examples of such tests are given.

4. Some Significance Tests for the Mean Using the Sample Range and Midrange. JOHN E. WALSH, Princeton University.

Consider a sample of size n, $(2 \leq n \leq 10)$, drawn from a normal population with mean μ. Let x_n be the largest value and x_l the smallest value of the sample. Significance tests are developed to compare μ with a given hypothetical value μ_0 by use of the sample. These significance tests are based on the quantity $D = [\frac{1}{2}(x_1 + x_n) - \mu_0]/(x_n - x_1) = [(\text{sample midrange}) - (\text{hypothetical mean})]/(\text{sample range})$. One-sided and symmetrical tests are considered. Values of D_α such that $Pr(D > D_\alpha \mid \mu = \mu_0) = \alpha$ are computed for $\alpha = .05, .025, .01, .005$. These values of D_α can be used to obtain one-sided tests at the .05, .025, .01, .005 significance levels and symmetrical tests at the .10, .05, .02, .01 significance levels. Efficiencies are computed for one-sided tests at the .05 and .01 significance levels. The efficiency is at least 90% for $n \leq 6$ at the .05 significance level and for $n \leq 8$ at the .01 level. The range-midrange test can be applied without computation through the use of an easily constructed graph. The application of a test requires only the plotting of the sample point (x_1 , x_n) on this graph.

5. Testing Compound Symmetry in a Normal Multivariate Distribution. DAVID F. VOTAW, JR., Princeton University.

Let $F(X)$ be the d.f. of a t-order vector variate $X(t \geq 3)$. Suppose the components of X are divided into mutually exclusive and exhaustive subsets. $F(X)$ is said to be *compound symmetric*, for the given division of its variates into subsets, if it is invariant over all permutations of its variates within these subsets. $F(X)$ is *completely symmetric* if the invariance holds over all permutations of its variates. If $F(X)$ is normal and compound symmetric, then within each subset of variates the means are equal, the variances are equal and the covariances are equal, and between any two subsets of variates the covariances are equal. Testing hypotheses of compound or complete symmetry in a normal $F(X)$ is of interest, for example, in studying psychological examinations and in medical research.

In this paper likelihood ratio criteria are developed for testing various hypotheses involving compound symmetry in regard to a normal distribution and to k normal distributions ($k \geq 2$). Given that the corresponding null hypothesis is true, the moments of each criterion are obtained explicitly and the distribution of each criterion is identified as the product of independent beta variates (in the case of a single normal distribution, the distributions are given explicitly for $t = 3, 4$, and 5 for certain divisions of the variates into subsets). In a previous paper Wilks has given results on a very thorough study of the sampling theory of likelihood ratio criteria for various hypotheses involving complete symmetry in regard to a normal distribution.

6. Effects of Non-Normality at High Significance Levels. HAROLD HOTELLING, University of North Carolina.

The effects of non-normality in the underlying population on the probabilities of *significance* by customary statistical tests are not well understood, in spite of numerous attacks, both mathematical and experimental, on the problem. Chung's recent proof that the distribution of the Student ratio t has in samples from an arbitrary population a distribution approaching normality for large samples tends to confirm the common idea that non-normality makes little difference if only the sample is fairly large, but this holds only for a fixed range of values of t while the sample number N increases. The tail areas beyond a deviation which increases with N in certain ways often behave quite differently than in sampling from a normal population. If p is the probability that $| t | > t_0$ in samples of N from a normal population and p' is the corresponding probability for another population, it is shown that $\lim_{N \to \infty} \left\{ \lim_{t_0 \to \infty} (p'/p) \right\}$ may be zero or infinite or may take any finite value, even when the non-normal distribution involved is of simple and realistic continuous forms. The conditions that this limit be unity are concerned only with the *shoulders* of the population histogram, and have nothing to do with its moments or its

behavior at infinity or at its mean. This suggests that caution should be used in applying familiar tests with high significance levels; that further calculations should be directed toward making this caution quantitatively definite; and that the use of sample moments or cumulants cannot lead to the most appropriate criterion of non-normality for this purpose.

7. On the Problem of Similar Regions. E. L. LEHMANN, University of California, Berkeley, and HENRY SCHEFFÉ, University of California, Los Angeles.

If $X = (X_1, \cdots, X_n)$ is a set of random variables with a joint probability density depending on a set of parameters $\theta = (\theta_1, \cdots, \theta_m)$, and if $T = (T_1, \cdots, T_m)$ is a set of sufficient statistics for θ, then Neyman (*Phil. Trans. Roy. Soc. London*, Vol. 236 (1937), pp. 333–380) has proved that a region w in the space of X is similar with respect to θ if it has the following structure: The intersections $w(t)$ of w with the surfaces $T = t$ have the property that the conditional probability of the sample point X falling into w given that $T = t$ does not depend on t. In the present paper a necessary and sufficient condition is found for the regions with the above structure to be the only similar regions. This condition is shown to be satisfied for a certain class K of probability densities which contains as special cases all densities for which the totality of similar regions has been previously determined. In particular the partial differential equations which Neyman (*Annals of Math. Stat.*, Vol. 12 (1941), pp. 46–76) assumed were satisfied in his solution of the problem of similar regions are solved and it is shown that any density satisfying these equations belongs to the above class K.

8. Fourth Degree Exponential Function. L. A. AROIAN and MARGUERITE DARKOW, Hunter College.

It is shown that the fourth degree exponential function is supported by the Bernoulli probability function and the hypergeometric probability function as well as being the function for which the method of moments is the *best* method according to the criterion of maximum likelihood. In the general situation six moments, at most, are needed. The function is classified into two general groups depending on symmetry or asymmetry and each case is divided again into unimodal and bimodal distributions. Examples show that the function is very successful in graduating the main Pearson types and the Gram-Charlier Type A frequency function. Various generalizations of the exponential function are indicated. In addition to its wide generality, the greatest practical advantage of the new system is the simplicity of the numerical calculations.

9. A General Weak Limit Theorem for Independent Distributions. P. L. HSU, University of North Carolina. (Read by title.)

For every positive integer n let there be n distribution functions $F_{n1}(x)$, $F_{n2}(x)$, $\cdots$, $F_{nn}(x)$. Assume that $\lim_{n\to\infty} \text{Max}_{1\leq j\leq n}\{1 - F_{nj}(x) + F_{nj}(-x)\} = 0$. Let $F(x)$ be the convolution $F_{n1}(x)*F_{n2}(x)* \cdots *F_{nn}(x)$. Let $\psi(t) = mit + \int_{-\infty}^{+\infty} [e^{itx} - 1 - itx/(1 + x^2)](1 + x^2)/x^2 \, dG(x)$, with $G(x) \uparrow$ and $G(\infty) - G(-\infty) < \infty$. Let $F(x)$ be the (infinitely divisible) distribution law having $\exp \psi(t)$ as its characteristic function. In order to have $\lim_{n\to\infty} F_n(x) = F(x)$ at every continuity point of $F(x)$, it is necessary and sufficient that the following relations hold at every $x > 0$ such that $\pm x$ are continuity points of $G(y)$:

$$\text{(I)} \quad \lim_{n\to\infty} \sum_{j=1}^{n} \int_{|y|>x} dF_{nj}(y) = \int_{|y|>x} ((1 + y^2)/y^2) \, dG(y),$$

(II) $\lim_{n\to\infty} \sum_{j=1}^{n} \left\{ \int_{|y|>x} y^2 \, dF_{nj}^{(y)} - \left(\int_{|y|<x} y \, dF_{nj}(y) \right)^2 \right\} = \int_{|y|<x} (1 + y^2) \, dG(y),$

(III) $\lim_{n\to\infty} \sum_{j=1}^{n} \int_{|y|<x} y \, dF_{nj}(y) = m + \int_{|y|<x} y \, dG(y) - \int_{|y|<x} (1/y) \, dG(y).$

10. On the Maximum Partial Sums of Sequences of Independent Random Variables. K. L. CHUNG, Princeton University.

The asymptotic behavior of the maximum partial sums of a sequence of independent random variables is studied in this paper. Two groups of new limit theorems are established under general conditions. The first group deals with theorems of the *weak* type. The limiting distribution of the maximum partial sums is obtained with an estimate of the remainder, thus improving a recent result of Erdös and Kac. Another estimate is obtained for a different domain of variation, which plays an essential role in the sequel. These results correspond to the sharper forms of the central limit theorem. In the second group, theorems of the *strong* type are obtained, giving precise lower bounds (in the sense of probability) for the maximum partial sums. These results form the exact counterpart to the general form of the law of the iterated logarithm, due to Feller, which give the precise upper bounds. A summary of the main results and methods has appeared in *Proc. Nat. Acad. of Sci.*, Vol. 33 (1947), pp. 132–136.

11. Some Results on the Distribution of Quadratic Forms From Gaussian Stochastic Processes. (Preliminary report). HERMAN RUBIN, Cowles Commission.

If one considers the estimation of the parameters of a Gaussian stochastic process whose elements are continuous functions from the functional values over a finite interval, one often finds that certain parameters can be estimated exactly, and certain parameters can not. This result often depends on the distribution of quadratic functionals whose arguments are elements of the stochastic process under consideration. In this paper, it is shown that the elements of a certain class of quadratic functionals have distributions concentrated at a point, and that the elements of a different class do not; in this latter case, the characteristic function is computed.

12. Some Significance Tests for the Median which are Valid under Very General Conditions. (Preliminary Report) JOHN E. WALSH, Princeton University. (Read by title.)

Consider n independent values drawn from populations necessarily satisfying only: 1) Each population has a unique median. 2) The median has the same value φ for each population. 3) Each population is symmetrical. 4) Each population is continuous. (It is to be emphasized that no two of the values are necessarily drawn from the same population.) Significance tests are derived for φ on the basis of 1)–4). These significance tests are based on order statistics of certain combinations of order statistics, each combination being either a single order statistic of the n values or one-half the sum of two order statistics. The tests are invariant under permutation of the n values and reasonably efficient if the values represent a sample from a normal population. The significance levels are of the form $r/2^n$, $(r = 1, \cdots, 2^n - 1)$. Each value of r can be obtained for some one-sided significance test. Thus any significance level can be closely approximated if n is large. The major disadvantage of these tests is the limited number of suitable significance levels available for small values of n. This disadvantage is partially eliminated by the development of tests which have a specified significance level if the values are a sample from a normal

population and a significance level bounded near this specified value if only 1)–4) necessarily hold. Results based on 1)–4) are applied to several well known statistical problems: Tests are obtained for the mean on the basis of a large number of independent values from populations having the mean but little else in common. Also generalized results are obtained for the Behrens-Fisher problem, quality control, slippage tests, the sign test and cases where some of the n values are dependent.

13. Loss of Information in t-tests with Unbalanced Samples. (Preliminary Report) JOHN E. WALSH, Princeton University. (Read by title.)

Consider two normal populations, the first with mean a_1 and variance σ_1^2, the second with mean a_2 and variance σ_2^2, while σ_1/σ_2 has a known value C. If the hypothesis $a_1 = a_2$ is to be tested by a t-test (one-sided or symmetrical) using n_1 sample values from the first population and n_2 values from the second population ($n_1 + n_2 = n$, fixed), it is shown that this experiment is most powerful when $n_1/n_2 = \sigma_1/\sigma_2$ (integer considerations neglected). The t-tests satisfying this condition will be referred to as balanced t-tests. Thus information will be lost by not using a balanced experiment. A quantitative measure of the information lost by using given values of n_1 and n_2 is determined by the total sample size m, ($m_1 + m_2 = m$), of the balanced t-test (same significance level) which has approximately the same power. Then $n - m$ sample values are wasted by using (n_1, n_2) rather than (m_1, m_2), i.e. only $100m/n\%$ of the information obtainable per observation is used by (n_1, n_2). A symmetrical t-test with significance level 2α has the same value of m as a one-sided t-test with significance level α. For one-sided t-tests with significance level α: $m \doteq \frac{1}{2}(B + \sqrt{B^2 - 8A})$, where $B = 2 + A + K_\alpha^2/2$, $A = (C + 1)^2[1 - K_\alpha^2/2(n - 2)][C^2/n_1 + 1/n_2]^{-1}$, and K_α is the standardized normal deviate exceeded with probability α. This approximation to m is valid for $m \geq 5$ if $\alpha = .05$, $m \geq 6$ if $\alpha = .025$, $m \geq 7$ if $\alpha = .01$, $m \geq 8$ if $\alpha = .005$. (A fractional value of m represents an interpolated measure of the sample size of the equivalent balanced experiment.)

14. Some Theorems on the Bernoullian Multiplicative Process. T. E. HARRIS Princeton University. (Read by title.)

A single entity may have j descendents with probability P_j, ($j = 0, 1, 2, \cdots$). Each *first generation* entity has then the same procreative probabilities, etc. Let

$$f(s) = p_0 + p_1 s + \cdots$$

If z_n is the number of entities in the nth generation, it is known that $P(z_n = j)$ is given by the coefficient of s^j in the nth iterate $f[f \cdots (f)] = f_n(s)$. Let $Ez_1 = x, 1 < x < \infty$. Conditions are given insuring that as $n \to \infty$ the cumulative distribution of the variate z_n/x^n approaches a limit-function which is absolutely continuous except for a possible single jump. Let $g(u)$ be the corresponding frequency function. If $f(s)$ is a polynomial of degree k, let $q = \log_x k/(\log_x k - 1)$. Otherwise, $q = 1$. Then $g(u) \cdot \exp\{u^{q+\epsilon}\}$ [is, is not] summable $(0, \infty)$ according as ϵ is [negative, positive]. Behavior of $g(u)$ near $u = 0$ is also considered. Special cases are considered where $g(u) = \text{constant} \cdot u^{1/m-1} \cdot e^{-u/m}$, m a positive integer. Maximum likelihood estimates for the parameters $p_0, p_1, \cdots$, and x are obtained as functions of n successive values $z_1, z_2, \cdots, z_n$. Consistency, in a certain sense, is proved. A specialized method is given for finding the moment-generating function of the variate N, the smallest value of n such that $z_n = 0$.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. George E. Albert has been appointed to an associate professorship at the University of Tennessee.

Dr. T. W. Anderson, Jr. has been promoted to an assistant professorship in the Department of Mathematical Statistics at Columbia University. He is on leave the first half of the 1947–48 academic year at the Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University as a Guggenheim Fellow. During the second half of the academic year he will be at Cambridge University.

Associate Professor Max Astrachan has been promoted to a full professorship at Antioch College, Yellow Springs, Ohio.

Associate Professor T. A. Bancroft, who has been at the University of Georgia, Athens, Georgia, is now with the Statistical Laboratory, Alabama Polytechnic Institute, Auburn, Alabama.

Dr. M. S. Bartlett of Cambridge University has been appointed as Professor of Mathematical Statistics at the University of Manchester, Manchester, England. The position is a newly created one. Professor Bartlett indicates that this position is believed to be the first official professorship in mathematical statistics in England.

Professor M. A. Brumbaugh has accepted a position with the Bristol Laboratories Inc., Syracuse 1, New York.

Dr. Donald A. Darling has been appointed Research Associate at Cornell University.

Professor D. B. DeLury of the Virginia Polytechnic Institute has accepted a position with the Ontario Research Foundation, 43 Queens Park, Toronto 5, Canada.

Professor Abel Gauthier of the University of Montreal has been appointed Head of the Institute of Mathematics and Assistant-Secretary of the Faculty of Science at that institution.

Dr. Casper Goffman, former assistant professor in the Mathematics Department, University of Kentucky, is now in the Mathematics Department, University of Oklahoma, Norman, Oklahoma.

Mr. Philip Hardy has returned to the General Electric Company at Warren, Ohio after serving at Wright Field.

Dr. Carl F. Kossack, who has been with the Navy Department in Washington, D. C. as an Air Intelligence Specialist, has accepted an associate professorship in the Department of Mathematics at Purdue University.

Mr. Frank Jones Massey, Jr. is now teaching in the Department of Mathematics, University of Maryland, College Park, Maryland.

Dr. William Burton Michael, who has been Lecturer in Mathematics, Psychology and Educational Psychology at the University of Southern California, has now accepted an assistant professorship in the Department of Psychology, Princeton University. He is also a member of the Research Department, College Entrance Examination Board at Princeton.

Mr. Bernard Ostle, a former teaching assistant, School of Business Administration, University of Minnesota, is now at Iowa State College, Ames, Iowa.

Mr. Maurice H. Quenouille, who was formerly with the Rothamsted Experimental Station, Harpendon, Herts, England, has accepted the position of Lecturer in Statistics, Marischal College, University of Aberdeen, Scotland.

Dr. James A. Rafferty left the Department of Pathology, University of Rochester in June and has been appointed Chief of the Department of Statistics, Air University, School of Aviation Medicine, Randolph Field, Texas.

Miss Mary Ann Savas has accepted a position with General Motors, Detroit, Michigan.

Professor George J. Stigler, formerly with Brown University, is now teaching in the Department of Economics, Columbia University, New York, New York.

Professor E. L. Welker has resigned an associate professorship in mathematics at the University of Illinois to become Associate in Mathematics in the Bureau of Medical Economic Research of the American Medical Association.

Mr. Sol M. Wezelman, who completed his master's degree in actuarial science at the University of Michigan in June, has accepted a position as Assistant Actuary in the North Dakota State Department of Insurance, Bismarck.

Dr. Bertram Yood has received his doctorate at Yale and is now on the staff at Cornell University.

Mr. Earl K. Yost, Jr. has accepted a position with the General Electric Co. at the Hanford Engineering Project, Richland, Washington.

Professor James G. Smith, of Princeton University, died at Princeton on November 28, 1946.

Beginning with the October issue, the quarterly journal *Mathematical Tables and Other Aids to Computation* will publish a new feature section, "Automatic Computing Machinery," designed to disseminate information and news on research and development in the field of high-speed automatic calculating machinery. Material should fall under the general headings of Bibliography, Technical Developments, Discussion (including correspondence), and News. Contributions to this section are invited and should be addressed to Dr. E. W. Cannon, Head of the Mathematics Group, Machine Development Laboratory, National Bureau of Standards, Washington, D. C.

Institute of Numerical Analysis Established

Plans have been completed for the establishment of one of the newest units of the National Bureau of Standards—the Institute of Numerical Analysis—at the

University of California at Los Angeles, according to an announcement by Dr. Edward U. Condon, Director of the Bureau.

One of the giant high-speed electronic computing machines, now under development by the Bureau of Standards, will be installed at the Institute when completed. Design specifications call for high memory capacity and automatically sequenced mathematical operations from start to finish at speeds attainable only with electronic equipment.

The Institute has two primary functions. The first is research in applied mathematics aimed at developing methods of analysis which will extend the use of the high-speed electronic computers. The second is to act as a service group for Western industries, research institutions, and government agencies. The service function will include not only the use of the machines for problem solving but also assistance in the formulation of problems in applied mathematics of the more complex and novel types. Service operations are to be initiated immediately, using the latest types of commercially available computing equipment.

The decision to locate the Institute at the University of California at Los Angeles was made after a nation-wide survey by the National Bureau of Standards. Centers in the East and Middle West were considered as well as the Far West, but Los Angeles, it was decided, offered the widest range of possibilities for an Institute of Numerical Analysis. Concentration of aircraft industries and the presence of several major scientific institutions were critical in the choice of Los Angeles.

Election of Fellows

The Board of Directors announced at the Yale Meeting the election of the following members as Fellows of the Institute: Theodore W. Anderson, Jr., Alexander C. Aitken, David H. Blackwell, Georges Darmois, Ragnar Frisch, Robert C. Geary, Frederick Mosteller, Gerhard Tintner, Charles P. Winsor and John Wishart.

New Members

The following persons have been elected to membership in the Institute
(June 1 to August 29, 1947)

Baldwin, Helen Mildred, B.S. (Cornell) Research Associate in Statistics, Atomic Energy Project, *215 Avenue C, Rochester 5, N. Y.*

Blunk, Paul M., A.B. Teaching asst. and grad. student, Univ. of Calif., *Box 532, Fair Oaks, Calif.*

Bowden, George Edwin, B.S. (Duke) Teaching asst., Math. Dept., White Hall, Cornell Univ., Ithaca, N. Y. ·

Bradley, Ralph Allan, M.A. (Queen's Univ.) Grad. student, Univ. North Carolina, *Wellington, Ontario, Canada.*

Burton, Kenneth John, Head of Statistics Section, British Employers' Confederation, *16 Rutherwyke Close, Ewell, Surrey, England.*

Carlson, Phillip G., Jr., A.M. (Columbia) *148 Cornell Street, Roslindale 31, Mass.*

Carol, Bernard, M.S.E. (Columbia) Graduate student at Columbia Univ., *15 West 96th Street, N. Y.*

Clark, Sidney B., B.A. (George Wash. Univ.) Statistician, Public Roads Administration, *2728 Porter St., N.W., Wash., 8, D. C.*

Danielson, Theresa, M.A. (Univ. of Ill.) Mathematician at Brookhaven National Laboratory, *3512 Cambridge Ave., New York 63, N. Y.*

Dineen, Russell D. F., B.A. (Univ. of Delaware) Graduate student at Univ. of Delaware, *1318 French Street, Wilmington, Delaware.*

Diver, M. L., M.E. (Purdue) Consulting Engineer, *P.O. Box 1016, San Antonio 6, Texas.*

Erasmus, Josias C., M.S.E. (Univ. of Stellenbasch, South Africa) Research Officer, Grootfontein College of Agriculture, Middelburg, C-P, South Africa.

Gottlieb, Morris J., Ph.D. (Wash. Univ., St. Louis) Member of the Institute for Advanced Study, Washington University, St. Louis, Mo.

Greenwood, Joseph Arthur, A.B. (Harvard) Student at Harvard University, *66 Oxford St., Cambridge 38, Mass.*

Gysbers, Jack C., M.A. (Univ. of Calif.) Teaching asst., Dept. of Math., Univ. of Calif., *2029 Berkeley Way, Berkeley 4, Calif.*

Haskind, Mina, B S. (Brooklyn College) Student at Brooklyn College, *763 Eastern Parkway, Brooklyn 13, New York.*

Hauser, Dr. Philip M., Ph.D. (Univ. of Chicago) University of Chicago, Chicago 37, Ill.

Hoyt, Cyril J., Ph.D. (Univ. of Minn.) Research Associate, Dept. of Education, University of Chicago, Chicago, Ill.

Kern, Enrique Roberto, First Assistant, Institute of Biometry, Univ. of Buenos Aires, Rivadavia 8854, Buenos Aires, Argentina.

Mark, Abraham M., Ph.D. (Cornell) Mathematics Department, Univ. of Wisconsin, Madison, Wisconsin.

Moss, George G., II, B.A. (St. John's College, Annapolis) Actuarial Statistician, Metropolitan Life, *2771 Morris Ave., N. Y. 58, N. Y.*

Phillips, Bernard E., A.M. (Columbia) *Box 147, Cathedral Station, New York 25, N. Y.*

Radvanyi, Laszlo, Ph.D. (Univ. of Heidelberg) Professor of Economics, National Univ. of Mexico, *Donato Guerra 1, desp. 207, Mexico, D. F.*

Richardson, John M., Ph.D. (Cornell) Member of Technical Staff, Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

Royston, Robert W., M.S. (Univ. of Mich.) Asst. Prof., Math. Dept., Wash. & Lee Univ., *117 W. Washington St., Lexington, Virginia.*

Savas, Mary A., A.B. (Univ. of Mich.) Student at Univ. of Mich., *524 E. Second St., Monroe, Mich.*

Shepard, David H., A.B. (Univ. of Mich.) Research Analyst, Army Security Agency, *505 Randolph Street, Falls Church, Virginia.*

Throdahl, Monte C., B.S. (Iowa State College) Research Chemist in Charge of Rubber Lab., Monsanto Chemical Co., Nitro, West Virginia.

Uchytil, Jan, Doctor of Science, Chief of Production Control Dep. in Central Federation of Czech. Industry, *Praha II, Prikopy 14, Czech.*

Vergara, Jose, Doctor of Engineering, (Madrid) Professor of Economics, Madrid, Chief of the Bureau of Statistics, Dept. of Agric., Madrid. *5262 S. Blackstone Ave., Chicago 37, Illinois.*

Wei, Dzung-shu, Ph.D. (Univ. of Iowa) Prof. and Head of Math. Dept., St. John's Univ., Shanghai, *129 East 10th St., New York 3, N. Y.*

Wolfson, Jacob, B.A. (New York College) Statistician, Social Security Adm., *845 Brunswick Road, Essex, Maryland.*

REPORT ON THE NEW HAVEN MEETING OF THE INSTITUTE

The Tenth Summer Meeting of the Institute of Mathematical Statistics was held at Yale University, New Haven, Connecticut, Tuesday, September 2 through Thursday, September 4, 1947. The meeting was held in conjunction with the summer meetings of the American Mathematical Society and the Mathematical Association of America. The following 150 members of the Institute attended the meeting:

C. B. Allendoerfer, R. L. Anderson, H. E. Arnold, L. A. Aroian, H. M. Bacon, J. L. Barnes, W. D. Baten, R. E. Bechhofer, A. A. Bennett, Joseph Berkson, D. H. Blackwell, C. I. Bliss, Colin Blyth, Jr., A. E. Brandt, G. M. Brown, R. H. Brown, O. P. Bruno, P. T. Bruyere, Mrs. P. T. Bruyere, J. H. Bushey, B. H. Camp, G. C. Campbell, Uttam Chand, K. L. Chung, W. G. Cochran, A. C. Cohen, Jr., E. P. Coleman, T. F. Cope, G. M. Cox, C. C. Craig, E. L. Crow, H. B. Curry, G. B. Dantzig, M. D. Darkow, B. B. Day, Bernard Dimsdale, C. E. Dieulefait, C. W. Dunnett, Churchill Eisenhart, L. R. Elveback, M. W. Eudey, H. P. Evans, William Feller, C. D. Ferris, M. M. Flood, R. M. Foster, H. A. Freeman, J. E. Freund, H. P. Geiringer, M. J. Gottlieb, J. Arthur Greenwood, Evelyn Groosman, F. E. Grubbs, H. T. Guard, P. R. Halmos, Max Halperin, M. H. Hansen, B. I. Hart, Mina Haskind, Wassily Hoeffding, R. H. Hoskins, Harold Hotelling, A. S. Householder, Jaroslav Janko, Irving Kaplansky, Leo Katz, Oscar Kempthorne, E. M. Kennedy, W. L. Kichline, C. J. Kirchen, L. F. Knudsen, H. S. Konijn, C. F. Kossack, Jack Laderman, H. G. Landau, E. L. Lehmann, R. A. Leibler, Walter Leighton, Jr., F. C. Leone, Joseph Lev, Howard Levene, Julius Leiblein, Arthur Linder, S. B. Littauer, E. D. Lowry, H. F. MacNeish, P. J. McCarthy, John Mandel, H. B. Mann, Sophie Marcuse, F. J. Massey, Margaret Merrell, E. B. Mode, M. E. Moore, Frederick Mosteller, D. N. Nanda, P. M. Neurath, M. C. Neurdenburg, G. E. Noether, M. L. Norden, H. W. Norton, P. S. Olmstead, A. L. O'Toole, E. R. Ott, T. E. Oxtoby, Edward Paulson, M. P. Peisakoff, G. B. Price, J. A. Rafferty, L. J. Reed, C. J. Rees, P. R. Rider, John Riordan, H. E. Robbins, Milton da Silva Rodrigues, A. C. Rosander, Ernest Rubin, Herman Rubin, Frank Saidel, M. M. Sandomire, Arthur Sard, Max Sasuly, F. E. Satterthwaite, E. D. Schell, Jack Sherman, Rosedith Sitgreaves, Andrew Sobczyk, Milton Sobel, Herbert Solomon, Mortimer Spiegelman, Arthur Stein, Henry Teicher, R. M. Thrall, Gerhard Tintner, M. N. Torrey, J. W. Tukey, D. F. Votaw, Jr., Abraham Wald, H. M. Walker, J. E. Walsh, R. M. Walter, J. H. Watkins, Dzung-shu Wei, E. S. Weiss, S. S. Wilks, C. P. Winsor, H. O. Wold, Jacob Wolfowitz, C. A. Wright, Bertram Yood.

The Tuesday afternoon session was devoted to a symposium on 2 x 2 tables with Professor Lowell J. Reed of Johns Hopkins University serving as chairman. Addresses were given on *Tests of Significance* by Dr. Churchill Eisenhart, National Bureau of Standards; *Estimation* by Dr. Charles P. Winsor, Johns Hopkins University and *Non-Standard Cases* by Dr. Joseph Berkson, Mayo Clinic. Discussants were Mr. William F. Taylor, Dr. Frederick Mosteller, Professor David H. Blackwell and Professor John W. Tukey. The attendance was approximately 130.

The first Wednesday morning session was devoted to contributed papers. Professor John W. Tukey of Princeton University presided. The attendance was approximately 85. The following three papers were presented:

1. *Estimation of Parameters in Truncated Pearson Frequency Distributions.*
   Professor A. C. Cohen, University of Georgia.

2. *Distribution of a Root of a Determinantal Equation.*
   Mr. D. N. Nanda, University of North Carolina.
3. *The Power of Certain Non-Parametric Tests of Independence.*
   Dr. Wassily Hoeffding, University of North Carolina.

The second Wednesday morning session was held with Professor Will Feller, President of the Institute, presiding. Professor R. A. Fisher, University of Cambridge, gave the address under the title *The Fitting of Gene Frequencies to Data for Genotypes.* The attendance was approximately 160.

The membership business meeting of the Institute was held at 9:15, Thursday morning, in 102 Chittenden Hall with President Feller presiding. The attendance was approximately 55. It was voted to make certain changes in the By-Laws and in particular to raise the due to $7.00 per year. (An exception is made for those living outside the Western Hemisphere.)• Morris Hansen, Chairman of the Committee on Planning and Development, initiated a lively discussion with reference to desirable changes in the Constitution.

On Thursday morning at 10:30, with President Feller presiding, Professor A. Wald of Columbia University presented the Henry Lewis Rietz Lecture on *Sequential Estimation and Multi-Decisions.* The attendance was approximately 150.

A joint session with the American Mathematical Society was held early Thursday afternoon at which Professor S. S. Wilks of Princeton University gave a lecture on *Sampling Theory of Order Statistics.* Professor Harold Hotelling of the University of North Carolina was the presiding officer. The attendance was approximately 300.

This session was followed by another joint session with the American Mathematical Society which was devoted to contributed papers. Professor John W. Tukey presided at this session and the attendance was approximately 115. The following seven papers were presented:

1. *Some Significance Tests for the Mean Using the Sample Range and Midrange.*
   Mr. John Walsh, Princeton University.
2. *Testing Compound Symmetry in a Normal Multivariate Distribution.*
   Dr. David F. Votaw, Jr., Princeton University.
3. *Effects of Non-Normality at High Significance Levels.* Professor Harold Hotelling, University of North Carolina.
4. *On the Problem of Similar Regions.*
   Dr. Erich L. Lehmann, University of California, Berkeley and Professor Henry Scheffe, University of California at Los Angeles.
5. *The Fourth Degree Exponential Function.*
   Dr. Leo A. Aroian and Professor Marguerite Darkow, Hunter College.
6. *On the Maximum Partial Sums of Sequences of Independent Distributions.*
   Dr. K. L. Chung, Princeton University.
7. *Some Results on the Distribution of Quadratic Forms from Gaussian Stochastic Processes.*
   Mr. Herman Rubin, Cowles Commission.
   The following four papers were presented by title:
8. *A General Weak Limit Theorem for Independent Distributions.*
   Professor P. L. Hsu, University of North Carolina.

9. *Some Significance Tests for the Median which are Valid under Very General Conditions* (Preliminary Report).
   Mr. John E. Walsh, Princeton University.
10. *Loss of Information in t-tests with Unbalanced Samples* (Preliminary Report).
    Mr. John E. Walsh, Princeton University.
11. *Some Theorems on the Bernoullian Multiplicative Process.*
    Mr. T. E. Harris, Princeton University.

Abstracts of all these papers appear elsewhere in this issue of the *Annals*.

A beer party in honor of the foreign statisticians attending the meeting was held in the dining room of Saybrook College on Wednesday evening. A joint dinner with the American Mathematical Society and the Mathematical Association of America was held on Thursday evening.

C. C. CRAIG,
Acting Secretary.

CONTENTS

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER, 1947

Articles

AMERICAN STATISTICAL ASSOCIATION
1603 K Street, N. W., Washington 6, D. C.

INDIAN AGRICULTURAL RESEARCH INSTITUTE LIBRARY, NEW DELHI.

Date of Issue	Date of Issue	Date of Issue